4

Causation with a Human Face

JIM WOODWARD

The recent literature on causation presents us with a striking puzzle. On the one hand, (1) there has been an explosion of seemingly fruitful work in philosophy, statistics, computer science, and psychology on causal inference, causal learning, causal judgment, and related topics.[1] More than ever before, 'causation' is a topic that is being systematically explored in many different disciplines. This reflects the apparent usefulness of causal thinking in many of the special sciences and in common sense. On the other hand, (2) many[2] philosophers of physics, from Russell onwards, have claimed that causal notions are absent from, or at least play no foundational role in, fundamental physics, and that at least some aspects of ordinary causal thinking (e.g. the asymmetry of the cause–effect relation) lack any sort of grounding in fundamental physical laws. If we also think that (3) if causal notions are appropriate and legitimate in common sense and the special sciences, then these notions must somehow reflect or derive from features of causal thinking (or true causal claims) that can be found in fundamental physics, then (1) and (2) appear to be (at the very least) in considerable tension with one another.

I will not try to systematically evaluate all of the many different claims made by philosophers about the role of causation in physics—the topic strikes me as very complex and unsettled and is anyway beyond my

Thanks to Chris Hitchcock, Francis Longworth, John Norton, Elliott Sober, and an anonymous referee for very helpful comments on an earlier draft and to Clark Glymour for extremely helpful correspondence in connection with Section 4.5.

[1] See, for example, Spirtes, Glymour, and Scheines (2000), Pearl (2000), and Gopnik and Schultz (forthcoming).

[2] Field (2003) and Norton (this volume), are among those expressing skepticism of one sort or another about the role of causal notions in physics.

competence. While I find the unqualified claim that causal notions are entirely absent from fundamental physics unconvincing (for reasons described on pp. 68–9), I am also inclined to think that there is something right in the claim that there are important differences between, on the one hand, the way in which causal notions figure in common sense and the special sciences and the empirical assumptions that underlie their application and, on the other hand, the ways in which these notions figure in physics. The causal notions and assumptions that figure in common sense and the special sciences do not always transfer smoothly or in an unproblematic way to all of the contexts in which fundamental physical theories are applied and common sense causal claims often do not have simple, straightforward physical counterparts. My aim in this paper is to explore some of the features of the systems studied by the upper-level sciences and the epistemic problems that they present to us that make the application of certain causal notions and patterns of reasoning seem particularly natural and appropriate. I will suggest that these features are absent from some of the systems studied in fundamental physics and that when this is so, this explains why causal notions and patterns of reasoning seem less appropriate when applied to such systems. There is thus a (partial) mismatch or failure of fit between, on the one hand, the way we think about and apply causal notions in the upper-level sciences and common sense and, on the other, the content of fundamental physical theories. Russell was right about the existence of this mismatch even if (as I believe) he was wrong in other respects about the role of causation in physics.

My general stance is pragmatic: the legitimacy of causal notions in the upper level sciences is not undermined by the disappearance (or non-applicability) of some aspects of these notions in fundamental physics. Instead, causal notions are legitimate in any context in which we can explain why they are useful, what work they are doing, and how their application is controlled by evidence. I thus reject claim (3) above: even if the claims of philosophers of physics about the unimportance or disutility of causal thinking in physics are correct, it still would be true that causal thinking is highly useful—indeed indispensable in other contexts.

My discussion is organized as follows. Section 4.1 comments briefly on the role of causation in physics. Section 4.2 discusses the role of causal reasoning in common sense and the special sciences. Section 4.3 sketches an interventionist account of causation that I believe fits the upper-level sciences and common sense causal reasoning better than competing approaches. Sections 4.4–4.8 then describe some of the distinctive features of the systems investigated in such sciences and way in which we reason about them. These include the fact that the causal generalizations we are able to formulate regarding the behavior of such systems describe relationships that are invariant only under a limited range of changes (Section 4.4), that such systems are described in a coarse-grained way (Section 4.5), that the systems themselves are located in a larger environment which serves as a potential source of 'exogenous' interventions (Section 4.6), that they have certain other features (including the possibility of 'arrow-breaking') that make the notion of an intervention applicable to them in a natural and straightforward way (Section 4.7), and that the epistemic problem of distinguishing causes and correlations, which arises very frequently in the special sciences, seems much less salient and pressing in fundamental physics (Section 4.8). Section 4.9 then explores the issue of the reality of macro-causal relationships in the light of the previous sections.

4.1 Causation in Physics

As I have said, my primary focus in this essay will be on the assumptions that guide the application of causal reasoning in common sense and the special sciences and make for its utility and how these differ from some of the assumptions that characterize the underlying physics. However, to guard against misunderstanding, let me say unequivocally that it is *not* part of my argument that causal notions play no role in or are entirely absent from fundamental physics. I see no reason to deny, for example, that forces cause accelerations. Indeed, as Smith (forthcoming) observes, there are numerous cases in which physics tells us how a local disturbance or intervention will propagate across space and time to affect the values of other variables, thus providing information that is 'causal' in both the interventionist sense described below and also in the sense captured by causal process theories such as Salmon (1984) and Dowe (2000).

It also seems uncontroversial that as a matter of descriptive fact various causally motivated conditions and constraints play important roles in physical reasoning and the application of physical theories to concrete situations. Thus advanced solutions of Maxwell's equations are commonly discarded on the grounds that they are 'non-causal' or 'causally anomalous' (Jackson 1999, Frisch 2000), candidates for boundary conditions involving non-zero fields at infinity or accelerations that are not due to forces may be rejected on the grounds that they are unphysical or acausal (Jackson 1999), various 'locality' conditions may be motivated by causal considerations and so on. However, it is also unclear exactly what this shows. Suppose (what is itself a disputed matter)[3] that in every case such causal constraints could be replaced by a more mathematically precise statement that does not use the word 'cause'. Would this demonstrate that causal notions play no fundamental role in physics or should we instead think of the replacement as still causal in content and/or motivation but simply more clear? Should we think of the constraints as in every case holding as a matter of fundamental law, which would allow us to say that whatever causal content a physical theory has is fixed by its fundamental laws, or should we instead retain the usual distinction between laws and boundary conditions, and (as seems to me more plausible and natural) hold that some of the causal content of the theory is built into assumptions about initial and boundary conditions, and is not carried by the laws alone? I will not try to resolve these questions here.

While causal claims and considerations are not absent from physics, certain commonly held philosophical assumptions about the role of such notions in physics and their connection to 'upper level' causal claims seem much more dubious. For example, in contrast to the assumption that all fundamental physical laws are causal (e.g. Armstrong 1997), many do not have a particularly causal flavor if only because of their highly abstract and schematic quality. Thus applications of the Schrödinger equation to particular sorts of systems in which a particular Hamiltonian is specified and specific assumptions about initial and boundary conditions adopted seem more 'causal' than the bare Schrödinger equation itself and similarly for many other examples.[4] When a fundamental physical theory is applied globally, to the entire universe, it is arguable that some of the presuppositions for the application of the notion of an intervention are not satisfied (cf. Section 4.6). Hence it may be unclear how to interpret what is going on in causal terms when 'causal' is understood along interventionist lines. A

[3] See Frisch (2002).

[4] See Smith (forthcoming) for more extended discussion.

similar point holds for some quantum mechanical contexts.[5] If, alternatively (or in addition) we think of causal claims as having to do with unfolding of causal processes in time from some local point of origin, as a number of philosophers (e.g. Salmon 1984, Dowe 2000) do, then few if any fundamental laws are causal in the sense of directly describing such processes[6] (Smith forthcoming). In part for these reasons and in part for other reasons that will be discussed on p. 73, the widely accepted idea that all true causal claims in common sense and the special sciences 'instantiate' fundamental physical laws which are causal in character, with causal status of the former being 'grounded' in these fundamental laws alone is deeply problematic.[7] In short, while there is no convincing argument for the conclusion that causation 'plays no role' in fundamental physics, there is also good reason to be skeptical of a sort of causal foundationalism or fundamentalism according to which fundamental physical laws supply a causal foundation for all of the causal claims occurring in the special sciences, and according to which, every application of a fundamental physical theory must be interpretable in terms of a notion of 'cause' possessing all of the features of the notion that figures in common sense and the special sciences.

4.2 Causation in the Special Sciences

Whatever may be said about the role of causation in physics, it seems uncontroversial that causal claims play a central role in many areas of human life and inquiry. Causal notions are of course ubiquitous in common sense reasoning and 'ordinary' discourse. Numerous psychological studies detail

[5] For example, it is arguable, (cf. Hausman and Woodward 1999) that there is no well-defined notion of an intervention on the spin state of one of the separated particles pairs with respect to the other in EPR type experiments. This would represent a limitation on the application of an interventionist account of causation only if there was reason to suppose that there is a direct causal connection between these states. Hausman and Woodward argue that there is no such reason; hence that it is a virtue, rather than a limitation in the interventionist account that, in contrast to other accounts of causation, it does not commit us to such a connection.

[6] Nor, contrary to 'conserved quantity' accounts of causation of the sort championed by Salmon and Dowe, will it always be possible to characterize 'causal processes' in fundamental physical contexts in terms of the transference of energy and momentum in accordance with a conservation law. Typically, the spacetimes characterized in General Relativity lack the symmetries that permit the formulation of global (that is, integral versions of) conservation laws. In such cases, no spatially/temporally extended process (no matter, how intuitively 'causal') will possess well-defined conserved energy/momentum. See Rueger (1998) for more detailed discussion.

[7] Davidson (1967) is a classic source for this idea.

the early emergence of causal judgment and inference among small children and the central role that causal claims play in planning and categorization among adults.[8] There is also considerable evidence that our greatly enlarged capacities for causal learning and understanding and the enhanced capacity for manipulation of the physical world these make possible are among the most important factors separating humans from other primates (Tomasello and Call 1997). However, for a variety of reasons, it seems to me misleading to think of causation as merely a 'folk' concept that is absent from 'mature' science. For one thing, many disciplines that are commonly regarded as 'scientific', including such 'upper-level' or so-called 'special science' disciplines as the social and behavioral sciences, medicine and biology, as well as many varieties of engineering traffic extensively in explicit causal claims. In disciplines such as economics, one finds explicit discussion of various concepts of causation (e.g. 'Granger' causation, in the sense of Granger 1998, and the contrasting manipulationist conception associated with writers like Haavelmo 1944, and Strotz and Wold 1960) and various tests for causation. In portions of statistics, in literature on experimental design, and in econometrics, one finds a great deal of self-conscious thinking about the difference between causal and merely correlational claims and about the evidence that is relevant to each sort of claim. Similarly, in brain imaging experiments it is common to find neurobiologists worrying that such experiments provide (at best) correlational rather than causal knowledge. (The latter requiring information about what would happen under experimental interventions, such as trans-cranial magnetic stimulation or from 'natural experiments' such as lesions.) Although there are of course exceptions, many of the most perceptive practitioners of these disciplines do not seem to doubt the utility of thinking about their subjects in causal terms.

4.3 An Interventionist Account of Causation

When issues arise about the coherence or legitimacy of some notion, it is often a useful heuristic to ask what the notion is intended to contrast with—what is it meant to exclude or rule out, what difference are we trying to mark when we use it? I believe that when 'cause' and cognate notions are used in the special sciences and in common sense contexts, the

[8] For discussion of these studies within an interventionist framework, see Woodward (forthcoming b)

relevant contrast is very often with 'mere' correlations or associations. In particular, the underlying problematic is something like this: an investigator has observed some relationship of correlation or association among two or more variables, X and Y. What the investigator wants to know is whether this relationship is of such a character that it might be exploited for purposes of manipulation and control: if the investigator were to manipulate X (in the right way), would the correlation between X and Y continue to hold, so that the manipulations of X are associated with corresponding changes in Y? Or is it instead the case that under manipulation of X there would be no corresponding changes in Y, so that the result of manipulating X is to disrupt the previously existing correlation between X and Y, with manipulation of X being ineffective as way of changing Y? In the former case, we think of X as causing Y, in the latter the connection between X and Y is non-causal, a mere correlation that arises in some other way—for example, because of the influence of some third variable Z or because the sample in which the correlation exists is in some way unrepresentative of the population of interest. Many epistemic problems in the special sciences fit this pattern such as the following.

(1) An investigator observes that students who attend private schools tend to score better on various measures of academic achievement than public school students. Is this because (a) attendance at private schools causes students to perform better? Or, alternatively, (b) is this a mere correlation arising because, for example, parents who send their children to private school care more about their children's academic achievement and this causes their children to have superior academic performance, independently of what sort of school the child goes to? For a parent or educational reformer the difference between (a) and (b) may be crucial—under (a) it may make sense to send the child to private school as a way of boosting performance, under (b) this would be pointless.[9]

(2) It is observed that patients who receive a drug or a surgical procedure are more likely to recover from a certain disease in comparison with those who do not receive the drug or procedure. Is this because the drug or procedure causes recovery or is it rather because, for

[9] For discussion, see Coleman and Hoffer (1987).

example, the drug or procedure has been administered preferentially to those who were more likely to recover, even in their absence?

(3) There is undoubtedly a systematic correlation between increases in the money supply and increases in the general price level. Is this because money causes prices, as monetarist economists claim, or is it instead the case that the causation runs from prices to money or that the correlation is due to the operation of some third variable? The answer to this question obviously has important implications for monetary policy (cf. Hoover 1988).

In each of the above examples, investigators are working with a notion of causation that is closely associated with the contrast between *effective* strategies and *ineffective* strategies in the sense of Cartwright (1979): the causal connections are those that ground effective strategies.[10] Even if we are firmly convinced that causal notions play no legitimate role in fundamental physics, it is hard to believe that there is no difference between drugs that cure cancer and those that are merely correlated with recovery or that some procedures, such as randomized trials, are not better than others, such as consulting the entrails of sheep, for assessing claims about the causal efficacy of such drugs. It is this general line of thought that motivates interventionist or manipulationist accounts of causation. For the purposes of this paper, I will describe a relatively generic version of such a theory—details are provided in Woodward (2003) and broadly similar views are described in Spirtes, Glymour, and Scheines (2000), and in Pearl (2000).

The basic idea is that causal claims are understood as claims about what would happen to the value of one variable under interventions on (idealized experimental manipulations of) one or more other variables. A simple statement of such a theory that appears to capture a type level notion of causal relevance, called total causation in Woodward (2003), is:

(C) X is a total cause of Y if and only if under an intervention that changes the value of X (with no other intervention occurring) there is an associated change in the value of Y.

[10] My claim at this point is simply that *in the examples under discussion*, the contrast between those relationships that are causal and those that are merely correlational coincides with the contrast between those that will or will not support manipulations. It is of course a further and more difficult question whether it is plausible and illuminating project to explicate the notion of causation in general along 'interventionist' lines, even when there is no practical possibility of manipulation. See Woodward (2003, ch. 3) for additional discussion.

Providing an appropriate characterization of the notion of an intervention is a matter of some delicacy. To see what is at issue, consider the following familiar causal structure, in which atmospheric pressure *A* is a common cause of *B*, the reading of a barometer, and *S*, a variable representing the occurrence/non-occurrence of a storm, with no causal link from *B* to *S* or vice-versa.

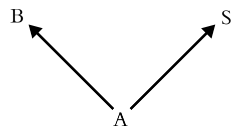

Figure 4.1

Clearly, if I manipulate *B* by changing the value of *A*, then the value of *S* will also change, even though, *ex hypothesi*, *B* is not a cause of *S*. If we want the relationship between behavior under manipulation and causation embodied in (C) to hold, we need to characterize the notion of an intervention in such a way as to exclude this sort of possibility. Intuitively, the idea that we want to capture is this: an intervention *I* on *X* with respect to a second variable *Y* causes a change in the value of *X* that is of such a character that *if* any change occurs in the value of *Y*, it occurs only as a result of the change in the value of *X* caused by *I* and not in any other way. In the case of the *ABS* system such an intervention might be carried out by, for example, employing a randomizing device whose operation is independent of *A* and, depending just on the output of this device, setting the position of the barometer dial to high or low, in a way that does depend on the value of *A*. If, under this operation, an association between *B* and *S* persists, we may conclude that *B* causes *S*; if the association disappears *B* does not cause *S*.

One way of making the notion of an intervention more precise, suggested by the above example, is to proceed negatively, by formulating conditions on interventions on *X* that exclude all of the other ways (in addition to *X*'s causing *Y*) in which changes in *Y* might be associated with a change in *X*. The following formulation, which is taken from Woodward (2003), attempts to do this.

Let *X* and *Y* be variables, with the different values of *X* and *Y* representing different and incompatible properties possessed by the unit *u*, the intent being to determine whether some intervention on *X* produces changes in *Y*. Then *I* is an intervention variable on *X* with respect to *Y* if, and only if, *I* meets the following conditions:

(IV)

I1. *I* causes *X*.

I2. *I* acts as a switch for all the other variables that cause *X*. That is, certain values of *I* are such that when *I* attains those values, *X* ceases to depend upon the values of other variables that cause *X* and instead only depends on the value taken by *I*.

I3. Any directed path from *I* to *Y* goes through *X*. That is *I* does not directly cause *Y* and is not a cause of any causes of *Y* that are distinct from *X* except, of course, for those causes of *Y*, if any, that are built into the *I*−*X*−*Y* connection itself; that is, except for (a) any causes of *Y* that are effects of *X* (i.e. variables that are causally between *X* and *Y*) and (b) any causes of *Y* that are between *I* and *X* and have no effect on *Y* independently of *X*.

I4. *I* is independent of any variable *Z* that causes *Y* and is on a directed path from *I* to *Y* that does not go through *X*.

I2 captures the idea that an intervention on *X* should place the value of *X* entirely under the control of the intervention variable *I* so that the causal links between *X* and any other cause *Z* of *X* are severed. This helps to ensure that any remaining association between *X* and *Y* will not be due to *Z*. For example, by making the value of *B* entirely dependent on the output of the randomizing device in the example above, we ensure that any remaining association between *B* and *S* will not be due to *A*. This idea lies behind the 'arrow breaking' conception of interventions described by Spirtes, Glymour and Scheines (2000), and Pearl (2000). I3 and I4 eliminate various other ways in addition to *X*'s causing *Y* in which *X* and *Y* might be correlated under an intervention on *X*.

One may think of the characterization IV as attempting to strip away, insofar as this is possible, the anthropocentric elements in the notion of an experimental manipulation—thus there is no explicit reference in IV to human beings and to what they can or cannot do and the characterization is given entirely in causal and correlational language.[11] In this respect,

[11] Because IV characterizes interventions in causal terms (the intervention causes a change in *X*, must bear a certain relationship to other causes of *Y* etc.), it follows that any account of causation like

the resulting theory is different from the agency theory of causation developed by Menzies and Price (1993, see also Price 1991) which makes explicit reference to human agency and to the experience of free action in characterizing what it is for a relationship to be causal. This is not to deny, however, that other facts about the sorts of creatures we are and the way in which we are located in the world (e.g. the fact that we ourselves are macroscopic systems with a particular interest in the behavior of other macroscopic systems with certain features) have played an important role in shaping our concept of causation. I shall return to this topic later.

4.4 Limited Invariance and Incompleteness

The connection between intervention and causation described by (C) is very weak: for X to cause Y all that is required is that there be some (single) intervention on X which is associated with a change in Y. Typically, we want to know much more than this: we want to know *which* interventions on X will change Y, and *how* they will change Y, and under what background circumstances. On the account that I favor this sort of information is naturally expressed by means of two notions: invariance and stability. Consider a candidate generalization, such as Hooke's law (H) $F = -kX$, where $X =$ the extension of the spring, $F =$ the restoring force it exerts, and k is a constant that is characteristic of a particular sort S of spring. One sort of intervention that might be performed in connection with (H) will involve setting the independent variable in (H)—the extension of a spring of sort S—to some value in a way that meets the conditions (IV). Suppose that it is true that for some range of such interventions the restoring force will indeed conform to (H). Then (H) is *invariant* under this range of interventions. More generally, I will say that a generalization is invariant (*simpliciter*) if and only if it is invariant under at least some interventions. Such invariance is, I claim, both necessary and sufficient for

C will not be reductionist, in the sense that it translates causal into non-causal claims. (My inclination, for what it is worth, is to think that no such reduction of the causal to the non-causal is possible). At the same time, however, I would argue that vicious circularity is avoided, since an interventionist elucidation (along the lines of C) of X causes Y does not presuppose information about whether there is a causal relationship between X and Y but rather information about *other* causal relationships or their absence—e.g. information about the existence of a causal relationship between I and X. For more on the issue of reduction, see Woodward (2003).

a generalization to describe a relationship that is exploitable for purposes of manipulation and control and hence for the relationship to qualify as causal. Typically, though, causal generalizations will be such that they will continue to hold not just under some interventions but also under at least some changes in *background conditions*, where background conditions have to do with variables which do not explicitly figure in the generalization in question. When a generalization has this feature I shall say that it is *stable* under the background conditions in question.[12] For example, taken as a characterization of some particular spring, (H) is likely to be stable under some range of changes in the temperature of the spring, the humidity of the surrounding air, and changes in the spatio-temporal location of the spring. I will suggest later, on pp. 78–80, that it is a feature of the generalizations that we call laws that they are stable under a 'large' or particularly 'important' range of changes in background conditions.

When thus characterized, both invariance and stability are clearly *relative* notions: a generalization can be invariant under some range of interventions and not under others and similarly for stability under background conditions. This will be the case for (H)—if we intervene to stretch the spring too much or if we change the background circumstances too much (e.g. if we heat the spring to a high temperature), the restoring force will no longer be linear. It would be a mistake, however, to take this to show that (H) does not describe a genuine causal relationship, in the sense of causal that an interventionist theory tries to capture. (H) is causal in the sense that it does not express a mere correlation between X and F but rather tells us how F would change under some range of interventions on X. It thus has the feature that we are taking to be central to causation—it describes a relationship that can be used for manipulation and control. However, the range of invariance of (H) (the conditions under which it correctly describes how F would change under interventions on X) as well as its range of stability are, intuitively, rather limited, at least in comparison with generalizations that describe fundamental laws of nature. Some philosophers contend that it is a mark of a genuinely fundamental law that it holds under all physically possible circumstances. Even if

[12] In Woodward (2003), I did not distinguish between invariance and stability in this way, but rather used 'invariance' to cover both notions. Since it is invariance under interventions rather than stability that is crucial for causal status, it seems clearer to distinguish the two notions. For additional discussion of this notion of stability, see Woodward (forthcoming a).

we weaken this requirement to accommodate the fact that, as currently formulated, many generalizations that are commonly described as laws (like Schrödinger's equation and the field equations of General Relativity) do, or may, break down under certain conditions, it remains true that these generalizations hold, to a close approximation, under a large or extensive range of conditions and, furthermore, that the conditions under which they do break down can be given a relatively simple and perspicuous theoretical characterization. By contrast, the conditions under which (H) will break down are both extensive and also sufficiently disparate that they resist any simple summary.

This limited invariance and stability of (H) goes hand in hand with its *incompleteness*: the restoring force exerted by a spring is contingent on many additional conditions besides those specified in (H)—conditions having to do both with the internal structure of the spring and with its environment. If (but only if) we have a spring for which these conditions happen in fact to be satisfied, (H) will correctly describe how its restoring force will change in response to interventions on its extension. In this sense, (H) describes a causal relationship that is merely *locally* invariant and stable or a relationship of contingent or *conditional* dependency. In these respects, (H) is paradigmatic of the sorts of causal relationships that we typically are able to establish and operate with in the special sciences and in common sense.[13]

There are a variety of different ways in which such merely locally invariant/stable relationships can arise. One of the simplest possibilities, illustrated by the case of the spring, is that the value of one variable Y depends on the value of a second variable X according to some relationship $Y = G(X)$ when some third variable B assumes some value or range of values $B = [b_1, \ldots, b_n]$ but this dependence disappears or changes radically in form when B assumes other values outside this range. If, as a contingent matter of fact, (4.1) B usually or always takes values within the range of stability for $Y = G(X)$—at any event, around here and right now or for cases which are of particular interest for us—*and* if (4.2) as long as B is within this range, intervening to change the value of X will not change the value of B to some value outside this range or otherwise disrupt the relationship $Y = G(X)$, then this relationship will be at least locally invariant.

[13] See Woodward (2003) for a more extended defense of this claim.

Note that requirement (4.2) is crucial for invariance. Both Hooke's law (H) and the relationship between the barometer reading *B* and the occurrence of the storm *S* will break down under the right conditions. What distinguishes these relationships is (4.2): all interventions that change the value of *B* will disrupt the *B*–*S* association but the corresponding claim is not true for (H). The distinction between the generalization describing the *B*–*S* relationship, which is not invariant under *any* interventions (although it is stable under some changes in background conditions), and (H), which is invariant under *some* but not all interventions and thus has a limited but non-empty range of invariance, is a distinction *within* the class of generalizations that fail to hold universally, under all circumstances. This makes the notion of invariance particularly suited for distinguishing between causal and merely correlational relationships in the special sciences, since we cannot appeal to notions like universality and exceptionlessness to make this distinction.

I have been comparing generalizations like (H), which are invariant under only a rather limited range of interventions and stable under a limited range of background conditions, with other generalizations, like laws of nature, with a 'more extensive' range of invariance/stability. Can we be more precise about the basis for such comparisons? In the case of macroscopic causal generalizations we are particularly interested in invariance and stability under changes that are not too infrequent or unlikely to occur, around here, right now, and less interested in what would happen under changes that are extremely unlikely or which seem 'farfetched'. Consider the generalization,

(s) Releasing a 5kg rock held at a height of 2 meters directly above an ordinary champagne glass with no interposed barrier will cause the glass to shatter.

(S) is certainly invariant under some range of interventions which consist of rock releasings. It is also stable under many possible changes in background conditions of a sort that commonly occur around here—changes in wind conditions, humidity, ordinary variations in the structure of the glass and so on. Of course there are other physically possible interventions/background conditions under which (S) is not invariant/stable. For example, it is physically possible that the rock might be deflected by a meteor on its downward path or vaporized by a blast of high energy radiation from space in such a way that the glass is left intact. If we wished to add conditions to (S) that would ensure that its antecedent was genuinely nomologically sufficient for the shattering of the glass, we would need to exclude these and many other possibilities. However, occurrences of the sort just described are extremely rare at least at present in our vicinity. We may formulate a relatively invariant and stable generalization (albeit one that falls short of providing a nomologically sufficient condition) if we ignore them. This is the strategy that is typically followed in the special sciences. From the point of view of agents who are interested in manipulation and control, this strategy makes a great deal of sense: an agent who wishes to shatter the glass can 'almost guarantee' this outcome by dropping the rock. Similarly for a central bank that wishes to lower inflation by restricting growth in the money supply, assuming that the relevant generalization is invariant and stable under changes that are likely to occur.

Whether occurrences like meteor strikes are likely to occur around here is, of course, a contingent matter, dependent not just on the fundamental laws governing nature but on initial and boundary conditions that might have been otherwise. This illustrates one respect in which causal generalizations in the special sciences often rely on, or presuppose, various sorts of contingent facts that are not guaranteed by fundamental laws alone and why we should not expect that the 'truth-maker' for (or underlying physical explanation of) the claim that a generalization like (S) is relatively invariant and stable to be located just in facts about underlying physical laws. Note, though, that although contingent, it is an 'objective' matter (*not* a matter of idiosyncratic individual taste and opinion) whether the sorts of disrupters of (S) just described are likely to occur around here.[14]

4.5 Coarse-Graining

There is another, related feature of the causal generalizations of the upper level sciences that contributes to their distinctive character: the variables in upper level causal theories are extremely *coarse grained* from the point of view of fundamental physics. This encompasses several different ideas. First, typically a number of different microstates, distinguishable from the

[14] Whether all grounds for the assessment of relative invariance are similarly objective is a difficult question that I will not try to resolve here. For further discussion, see Woodward (2003) on 'serious possibility' and 'farfetchedness'.

point of view of fundamental physics, will realize the same value of the macro-variables of upper level causal theories. The relationship between thermodynamics and statistical mechanics (from which I take the notion of coarse-graining) is paradigmatic: if we think of a microstate as a complete specification of the position and momentum of each of the component molecules making up a sample of gas, then a very large number of such micro-states will realize a single value for the pressure and temperature of the gas. Similarly, a large number of different molecular states will constitute different ways of realizing the state or event which we describe as the shattering of a glass or the attainment of a certain number of years of schooling.

Second, while initial and boundary conditions in fundamental physics are often described by expressions specifying the exact values of variables at *each* space time point within some region, as when charge densities and field strengths are specified as a function of position in classical electromagnetism, causal generalizations in the special sciences often relate variables such that a single value of these characterizes an entire macroscopically spatially extended and perhaps temporally extended region, with boundaries that are not very precise from the point of view of fundamental physics. Typically, these regions and the objects associated with them, will be (from a macroscopic perspective) relatively connected and cohesive, or at least not too diffuse, discontinuous and gerry-mandered. As illustrations, consider generalizations relating the impact of a rock on a window and its subsequent shattering and position in a primate dominance hierarchy to level of serotonin expression. (high position in the hierarchy causes high levels of expression and vice-versa). The variables involved in these generalizations {*shattering, non-shattering*}, {*window breaks, window does not break*}, {*position in the dominance hierarchy is such and such*} and so on, take their values across extended spatio-temporal regions with imprecise boundaries. Thus the shattering occurs in the region of the window but beyond a certain level of discrimination, there may be no definite answer to exactly where this event is located or when it begins and ends. In addition, coarse grained variables may fail to completely partition the full possibility space as seen from the point of view of an underlying fine-grained theory.[15] Both episodes of shattering and of non-shattering are likely to be

[15] Thanks to Chris Hitchcock for emphasizing this point.

identified by means of their similarity to prototypical or paradigm cases. From a macroscopic perspective the possibility of an outcome intermediate between shattering and not shattering may not be recognized, even if this corresponds to some possible, albeit very unlikely microstate—e.g., the rock grazes the glass in such a way to crack it extensively and displace small portions of it while leaving most of the pieces in (almost) their previous position, so that the result is neither a prototypical shattering or non-shattering. To the extent that all this is so, it may also be unclear exactly which microstates are to be identified with or are realizers of the shattering.[16] However, to the extent that we are likely to be interested only in the contrast between the window's shattering and its not shattering at all, this indeterminacy will not matter much.

Finally, in common sense and the upper-level sciences, causal relata are often described as operating across spatiotemporal gaps or, alternatively, in a way that is non-specific about the spatiotemporal relationship between cause and effect. Recovery from a disease will typically occur some significant lapse of time after the administration of the drug that causes recovery. A slowdown in economic activity may be caused by the decision of the central bank to raise interest rates but it seems doubtful that there is any clear sense in which the latter event is spatiotemporally contiguous with the former. It is true that in many, but by no means all,[17] cases involving macro-causality, there will exist (from a more fine-grained perspective) a spatio-temporally continuous process linking the cause to its effect. However, even when such processes do exist, upper level causal generalizations often do not specify them and the correctness and utility of the upper level generalizations do not rest on our actually having information about such processes. This feature is captured nicely by interventionist accounts which take the distinctive feature of causal relationships to be exploitability for purposes of manipulation, regardless of whether there is a spatiotemporal gap between cause and effect.

[16] This is one reason, among many, why it is unsatisfactory to say, as Davidson does, that when rock impacts cause glass shatterings, both of these macroscopic events 'instantiate' an underlying law relating micrsostates.

[17] 'By no means all' because in some cases the underlying processes will involve causation by omission or by disconnection in the sense of Schaffer (2000) or some complex combination of these and spatio-temporally continuous processes. Presumably most cases of causation by omission and disconnection do not involve the direct instantiation of fundamental causal laws.

How do these features of causal generalizations in the upper level sciences compare with the laws of fundamental physics? In contrast to the incomplete relationships of limited invariance between coarse-grained factors that are characteristic of the upper level sciences, fundamental laws typically take the form of differential equations, deterministically relating quantities and their space and time derivatives at single spatiotemporal locations. In these equations, as Hartry Field has recently observed (2003), there are no spatial or temporal gaps of a sort that would allow for the possibility of outside influences intervening between the instantiation of the independent and dependent variables. In contrast to the imprecise spatiotemporal boundaries of the causal factors that are of interest in the upper-level sciences, the initial and boundary conditions required for solution of the equations of fundamental physics are described by specifying the exact values of the relevant variables at every spacetime point within some extended region. While the causal generalizations of the upper-level sciences are invariant and stable only under some limited range of changes, the ideal within fundamental physics is to find generalizations that are invariant and stable under all possible changes or, failing that, generalizations that break down only under well-specified extreme conditions and whose range of invariance and stability is thus far less limited than the generalizations of the special sciences. In other words, the ideal is to find generalizations that are complete (or nearly so) in the sense of incorporating independent variables describing, as nearly as possible, *all* quantities such that *any* variation in their values would lead to different values for the dependent variable in that generalization and which, taken together, are nomologically sufficient for (that is, physically guarantee, whatever else may happen) the value of the dependent variable. Again, this contrasts with the causal generalizations of the special sciences which are radically incomplete and fall well short of specifying such nomologically sufficient conditions.

As a number of writers have argued, to specify a set of conditions *S* that are genuinely nomologically sufficient for some event of interest *E*, we need (at least) a description of a cross section of the entire backward light cone for *E*—a description that specifies the values of the relevant variables at every point within this cross section. Anything less than this will leave open the possibility that the conditions *S* are satisfied and yet some influence compatible with *S* occurs which would exclude the occurrence of *E*. (A burst of high energy radiation from outer space that vaporizes the rock just before it strikes the glass etc.) If one mark of a fundamental law is that it describes such a nomologically sufficient condition, then it seems highly unlikely that there will be fundamental laws relating the localized, coarse-grained events in which the upper level sciences traffic.

In fact, if *E* is given a sufficiently precise microphysical description and the causes of *E* are also given a similar description and are taken to include every variable, some variations in the value of which would lead to the non-occurrence of *E*, then an even more disturbing consequence appears to follow—everything in the backward light cone of *E* will qualify as a cause of *E*. Suppose that *E* is the event of my headache disappearing at *t* and that in the absolute past of this event are (*A*) my ingestion of aspirin 30 minutes prior to *t*, and at about the same time, my next-door neighbor's sneezing (*S*) and my wishing (*W*) my headache would go away. Consider very fine-grained specifications ((E^*), (A^*), (S^*), and (W^*)) of the exact position and momentum of the fundamental particles realizing (*E*), (*A*), (*S*), and (*W*). The occurrence of *A* will alter the gravitational and electromagnetic forces incident on *E* (and hence E^*) but this will also be true of (*S*) and (*W*)—indeed any small variation in S^* and W^* will alter E^*, albeit in small ways. So will small changes in the gravitational influence of the distant stars. If we want a genuinely nomologically sufficient condition for E^* it looks as though we need an exact specification not just of A^*, but of S^*, W^* and much more besides.

As Field (2003) emphasizes, this is not just the unthreatening point that many other factors besides those that are salient to common sense are among the causes of *E*—that in ordinary discourse we pick out just a small part of the complete 'Millian' cause of *E*. (As when we say that the complete cause of the fire includes the presence of oxygen and the absence of a sprinkler system etc. as well as the more salient striking of the match.) This point is acknowledged by many theories of causation. Rather, the observation in the previous paragraph threatens to collapse the distinction between causal and temporal priority and with it the whole point of the former notion. Barring extraordinary circumstances we think that taking aspirin is (or at least may be) an effective strategy for making a headache go away and that sneezing and wishing are not. If we are forced to the conclusion that the latter are causes of *E* as well, the motivation (according to manipulationist accounts) for introducing the notion of causation in the first place—the intuitive connection between causation and manipulation and the contrast between causal and merely correlational relationships—appears lost.

Obviously the description of causes in ordinary life and in the special sciences does not take the form of a complete description of the values of relevant variables at every point on a surface intersecting the backwards light cones of effects of interest. The key to understanding how it is possible to provide genuine causal information without providing such a detailed micro-description can be found in the notions of coarse-graining and limited invariance/stability described above. One consequence of coarse-graining is that it makes it permissible to ignore certain causal factors that would be relevant at more fine-grained level of description. If the effect of interest is the exact position and momentum of some collection of particles, then all forces incident on these particles are causally relevant. Suppose, however, the effect variable is framed in a much more coarse-grained or chunkier way—for example, as recovery/non-recovery from a particular condition or disease. If our task is to find the variables, variations in the value of which account for the contrast between those experimental subjects who recover and others who do not, it will almost certainly no longer be true that everything in the backward light cone is relevant to this contrast. For example, it is extremely unlikely that any actual variations in the gravitational force exerted by the distant stars on the subjects will have anything to do with the contrast between recovery and non-recovery (either for a single subject or collection of subjects). Similarly, for whether the subjects or their neighbors sneeze, for what they wish for at the time they ingest medication, and for the color of the clothes they wear. Note that this is true even though, as already emphasized, each token event of sneezing, wishing and so forth will causally influence (in some respects) each of the micro-events that realize individual tokens of recovery or non-recovery. This is possible because, although the forces arising from, for example, the occurrence of a neighbor's sneeze will certainly influence the microstate of the subjects, they will not change these states sufficiently to turn a subject who otherwise would have recovered into a non-recoverer or vice-versa.[18]

[18] The choice of grain associated with the causal analysis of a situation is intimately related to the contrastive character of causal claims. As we alter the grain, we alter the potential contrastive foci that are available. If we employ a very fine-grained description of a shattered glass, we can ask why the shattered pieces are in exactly this particular configuration, rather than (or in contrast to) a slightly

To put the point more generally: not all causal relationships (or relationships of nomological dependency) among micro-events aggregate up to causal relationships among coarse-grained macro-events that are constituted by those micro-events. Instead, whether one gets causation at the macroscopic level will depend (among other things) on the particular coarse-graining that is chosen. Of course, we look for sets of coarse-grained variables which are such that not every variable in the set turns out to be causally related to every other, since under this possibility, the discovery of causal relations among macro-events would lose most of its interest and point. Although coarse-graining may look imprecise and arbitrarily selective from the point of view of the underlying physics, it makes the task of finding and describing causes much easier.

It may seem surprising, even counterintuitive, that causal and statistical dependence relationships involving fine grained microscopic variables do not automatically show up in causal and dependence relationships among the macroscopic variables that are realized by the fine grained variables. As we have seen, 'realization' is a fuzzy notion, but one simple framework for thinking about at least some cases of this sort is as follows:[19] Suppose that $Z_1, Z_2, \ldots, Z_n$ are fine-grained variables that stand in various statistical independence, dependence and conditional relationships to one another as represented by a joint probability distribution $P(Z_1, Z_2, \ldots, Z_n)$. Suppose that $F_1, F_2 \ldots F_n$ are functions that map, respectively, $Z_1, Z_2, \ldots, Z_n$, into more macroscopic variables, $F_1(Z_1), \ldots, F_n(Z_n)$. What can be said about

different configuration—indeed, the use of a fine-grained level of description naturally suggests this as an appropriate explanatory question and that the explanandum should be understood in terms of this contrastive focus. If, instead, we employ a much more coarse-grained description, according to which the only two possibilities are that the glass either shatters or does not, then only a very different contrastive focus is possible—we now ask why the glass shattered rather than not shattering at all. Obviously a differentiating factor that is relevant to the explanation of this second contrastive focus—e.g. that the rock struck the glass (rather than missing it entirely) explains the contrast between shattering and not shattering—may not be (in this case, is not) relevant to the explanation of the first, more fine-grained contrast. I would thus reject the anonymous referee's suggestion that coarse-graining and contrastivity are different, competing ways of understanding the cases under discussion. On my view, they are complementary and closely associated.

[19] Thanks to Chris Hitchcock for a helpful discussion and to Clark Glymour for some very helpful correspondence. The framework that follows corresponds to one very simple possibility, with each fine-grained variable being mapped directly into a macroscopic variable. Of course, there are many other possible relationships between the fine-grained and the macroscopic. For example, the macroscopic variable might be a sum or some other function of the values of some fine-grained variable taken by each of a large number of units. For some relevant results in this connection, see Chu, Glymour, Scheines, and Spirtes (2003).

how the (in)dependence relationships among these variables depend on the functions F_i, and the distribution P? It is not clear that there is any illuminating general answer to this question but here are some relevant observations that will help to motivate the claims made in the previous paragraphs. First, if Z_j and Z_k are unconditionally independent, then $F_j(Z_j)$ and $F_k(Z_k)$ will also be independent as long as F_j and F_k are measurable functions.[20] Second, if Z_j and Z_k are dependent, then $F_j(Z_j)$ and $F_k(Z_k)$ will be dependent if F_j and F_k are 1–1 functions. In addition, if the functions F_i are 1–1, they will preserve conditional independence (screening off) relationships among the Z_i (Glymour forthcoming). If F_j and F_k are not 1–1, then simple examples show that depending on the details of the case, it is possible to have Z_j and Z_k dependent but $F(Z_j)$ and $F(Z_k)$ independent.[21] Functions that are not 1–1 can also fail to preserve conditional independence relationships. If we think of coarse-graining as involving the use of functions from micro to macro-variables that are not 1–1 (different values of the micro-variable are mapped into the same value of the macro-variable) then coarse-graining can indeed lead from dependence at the micro-level to independence at the macro-level.

Observing that it is possible for macro-independence to emerge from micro-dependence is of course not at all the same thing as providing an interesting characterization of the conditions under which this will happen. In my view, one of the major puzzles about the relationship between causation at the macro-level and the underlying physics is why there is so *much* independence (or at least near or apparent independence) and conditional independence among macroscopic variables, given that at a microphysical level, everything seems so interconnected. Macroscopic independence is one of the conditions that allows us to discover and exploit macroscopic causal relations. Somehow this independence must result from coarse-graining and its interaction with facts about the underlying

[20] Glymour, personal correspondence.

[21] Two examples, the first my own and the second due to Hitchcock: First, suppose that X can take any of the values $\{1, 2, 3, 4\}$ and that the associated values of Y are respectively $\{1, 0, 1, 0)$ so that Y is a deterministic function of X. Suppose that each value of X is equally probable and when $X = 1, X = 2, F(X) = 1$ and when $X = 3, X = 4, F(X) = 0$. Let $G(Y) = Y$ be identity. Then $F(X)$ will be independent of $G(Y)$. Second, suppose that X is vector valued: $X = (X_1, X_2)$ and that Y and X are dependent because Y and X_1 are dependent, while Y is independent of X_2. Let $F(X) = X_2$, and let G be identity. Then $F(X)$ will be independent of $G(Y)$.

dynamics, but I would be the first to acknowledge that I have said very little about how this happens.[22]

Somewhat surprisingly, the coarse-grained character of the variables figuring in the incomplete causal generalizations of the special sciences enables those generalizations to be more stable than they would be if they remained incomplete but were formulated in terms of more fine-grained variables. In fact, there are few incomplete dependency relationships relating fine-grained variables that are stable over some usefully extensive range of changes in background conditions; instead all or most incomplete dependency relationships that are stable over even a modest range of changes will relate coarse-grained variables. Finding generalizations describing dependency relationships involving fine-grained variables that are relatively stable will typically require finding generalizations that are complete or nearly so.

Suppose that the state F at time t of some macroscopically spatially extended region R of the world is specified by a conjunction of fine-grained properties $P_1 \ldots P_n = F$ and that the state of the world F' at some previous time t' which is nomologically sufficient for Rs being F at t is given by the fine-grained conjunction $P'_1 \ldots P'_n$ Then I claim that in most cases and for most specifications of $P'_1 \ldots P'_n$ and $P_1 \ldots P_n$, we are unlikely to find even relatively stable generalizations of form: 'If P'_i, then P_i'. That is, generalizations that relate just one of the properties P'_i to just one of the properties P_i are likely to be highly non-invariant and exception-ridden. A similar point will hold if the generalizations in question relate a conjunction of some (but not all) of the P'_i to some Pi. The reason for this is that most fine-grained properties P_i of R at t will generally depend not on some single P'_i but rather on the entire specification $F' \ldots$—the entire conjunction $P'_1 \ldots P'_n$. Changing any one of the variables P'_k, $k \neq i$ (even slightly) will disrupt the relationship between P'_i and P_i. On the other hand, if C is some property holding in R at t that is sufficiently coarse-grained, it may well be true that we can find some coarse grained property C' holding at t' such that the generalization '*if* C', then C' is incomplete but relatively invariant and stable, over locally prevailing background conditions.

As an illustration, consider again an experiment in which an irregularly shaped, 5kg rock is dropped from a height of two meters directly onto an intact champagne glass, causing it to shatter. Suppose first that we specify

[22] See Strevens (2003) for interesting additional discussion of this issue.

this effect in a very fine-grained way: we take the effect to be the exact shape and location of all of the various fragments of the glass five seconds after its breaking. Let M_1 be such a specification for one of the fragments of the glass. Assume that the interaction is deterministic, and hence that there is some extremely complicated specification S of the state of the rock, the state of the glass, and the surrounding environment at the time of collision that is nomologically sufficient for M_1. Note that for S to be nomologically sufficient for M_1, S must include a specification not just of the momentum p of the rock, but also the shape of that portion of the rock that comes in contact with the glass, the composition and structure of the glass, and much more besides. If we try to formulate an incomplete generalization linking some incomplete specification of S (e.g. if we just specify the momentum p of the rock and nothing more) to M_1, this generalization will almost certainly be exception-ridden and relatively unstable. Given rocks that share the same momentum p, but differ in shape, orientation or larger environment when they strike the glass and so on, even slight changes in these variables will result in a set of fragments that are at least slightly different and similarly for slight changes in the composition, structure and so on of the glass.

Suppose instead that we adopt a much more coarse-grained description of this situation: we represent the effect by means of a variable Y that can take just two values: {1 = *glass shatters*, 0 = *does not shatter*} and the cause by means of a variable X that takes the following two values. {1 = *5kg rock is dropped from a height of 2 meters directly striking glass*, *no rock strikes glass*}. In most non-extraordinary background circumstances, an intervention that sets $X = 1$ will be reliably followed by $Y = 1$: that is, the generalization $Y = X$ describes a relatively stable dependency relationship. This illustrates how coarse-graining allows the formulation of incomplete generalizations that are relatively invariant, although at the cost of predictive precision regarding fine-grained details.

Although I lack the space for detailed discussion, it is worth noting that the use of coarse-grained variables affects many other aspects of causal reasoning in common sense and the special sciences. Consider the familiar screening-off conditions connecting causal claims and probabilities and the supposed fork asymmetry associated with these: we expect that common causes will screen-off their joint effects from one another but that conditioning on the joint effect of two causes will render them dependent except for certain very special parameter values. As Arntzenius (1990) notes, under determinism, if C, which occurs prior to the joint effects E_1 and E_2, screens them off from one another, there must also be an event C^* which occurs after E_1 and E_2 is causally affected by them, and that screens them off from one another. Why then are we tempted to suppose that there is a fork asymmetry? At least part of the reason is that typically the later screening-off event C^* will be very hard to see—it will be very diffuse, spread out and gerry-mandered, corresponding to no single macro-event in any coarse-graining we are likely to adopt. If, as we tacitly assume, the screening-off common effect must be a single event in a natural coarse-graining, then it becomes more plausible that there is a fork asymmetry. In this sense, the asymmetry is in part a product of the particular coarse-graining of the macroscopic world that we adopt. A similar story can be told, I believe, about how it is possible for the coarse-grained entropy of a system to increase despite the fact that fine-grained entropy is constant over time. In the paradigmatic case of a gas expanding into an evacuated chamber, we must coarse-grain in order to have an expansion of the phase volume with time, so that entropy increases. The expansion is irreversible relative to an appropriately chosen coarse-graining but not in relation to a fine-grained level of description.[23]

4.6 Interventions

I turn now to a different feature of the systems that are studied in the upper-level sciences and our relation to them that makes causal notions seem particularly useful or appropriate. This has to do with the fact that such systems are typically only a small part of a much larger world or environment which is outside the scope of the inquirer's interest but which can serve as source of interventions. Recall the basic idea of an interventionist account

[23] For reasons that are lucidly explained in Sklar (1993, pp. 346 ff), I believe that it does not follow from the fact that whether entropy increases is relative to the grain one chooses that whether or not entropy increase occurs is 'subjective'. Coarse-grained entropy is a different quantity than fine-grained entropy and these quantities behave differently. Relative to a specification of system and a level of description or graining for it, it is an objective matter whether there is entropy increase. A similar point holds for causation—once one fixes the variables one is talking about, it is 'objective' matter whether and how they are causally related. Causation is, one might say, 'variable relative' in the sense that, as illustrated, different choices of variables or grainings will lead to different conclusions about whether everything in the backward light cone is causally relevant to an episode of glass shattering but it is not 'description-relative' in the sense that whether or not one variable is causally relevant to another depends on how those variables are described. See Woodward (2003) for additional discussion.

of causation: causal claims are linked to counterfactual claims about what would happen under possible interventions. Clearly if such a view is to be remotely plausible, 'possible' must be understood in a liberal way. For example, there are many cases of causal relationships between X and Y in which an intervention on X is not practically or technologically possible for human beings. Woodward (2003) discusses this issue and concludes that for an interventionist account of what it is for X to cause Y to be workable, what is crucial is that counterfactuals describing what would happen to Y (or in the indeterministic case the probability distribution of Y) under an intervention on X 'make sense' and 'have determinate truth values', rather than whether human beings are able to carry out the interventions in question. But under what circumstances will interventionist counterfactuals have the quoted features?

While I will not try to provide a definitive answer to this question (and in fact doubt that there is an uncontroversially correct answer that covers all possibilities), I want to suggest that there are some circumstances and systems for which interventionist counterfactuals seem straightforward and unproblematic and other systems for which this is less obviously the case. Unsurprisingly, among the former systems are those investigated in the special sciences. The latter include global applications of fundamental physical theories to the whole universe or large portions of it. I emphasize again that I do *not* mean to claim that the notion of an intervention has no application in physics; the notion seems perfectly reasonable when applied to the right sort of small, non-global systems. I maintain, however, that the requirements for the sensible application of the notion of an intervention help to explain why forms of causal thinking that seem natural in the special sciences do not straightforwardly extend to more global physical contexts.

Consider a typical case in which interventionist counterfactuals seem unproblematic. A researcher wishes to know whether treatment with a drug D will cause an increase in the rate of the recovery from a certain disease. She envisions the following experiment. Subjects with the disease are randomly assigned on the basis of the outcome of the flip of a fair coin to a treatment group who receive D and a control group from whom D is withheld. Which subjects in the trial receive the drug is thus determined entirely by the random assignment process. Then the incidence of recovery in the treatment and control group is compared. In an experiment of this design it usually will be reasonable for the experimenter to assume that the random assignment process constitutes an intervention on who receives the drug with respect to the outcome of recovery. Of course this is a defensible empirical assumption that might in principle be mistaken—perhaps unbeknownst to the first researcher a second scientist controls the outcome of the coin toss with a magnet and arranges that all and only those with unusually strong immune systems are assigned to the treatment group. Usually, however, the assumption that this operation constitutes an intervention will be correct–whatever variables influence the outcome of the coin toss will not causally influence or be correlated with whether the subjects recover except via the route, if any, that goes from the outcome of the flip to ingestion of the drug to recovery. Even if the researcher does not in fact perform this experiment, it seems clear enough what would be involved in performing it and no reason to doubt that there is a determinate answer to the question of what would happen if it were to be performed.

One reason why interventionist counterfactuals seem unproblematic in this case is that we are dealing with what Judea Pearl (2000) has called a 'small world'—a system (the subjects in the experiment who are or are not given the drug) that is isolated enough from its environment that it can serve as a distinctive subject of causal inquiry but not so isolated (or 'closed') that the idea of outside influences in the form of interventions makes doubtful sense. Put slightly differently, the system of interest is located in a larger environment which serves as a potential source of 'exogenous' interventions. However, apart from this, the environment is of no direct interest to the researcher. In the example under discussion, the outcomes of the coin flip are exogenous and part of the environment in this sense: the researcher is not interested in, and does not need to be concerned with, modeling in detail the causal processes that produce these outcomes as long as it is true that these processes, whatever they may be, do not affect recovery, independently of treatment. In this sort of case, we can keep any contra-nomic miracles that the occurrence of interventions may seem to require safely offstage, in the environment.[24]

As several writers have remarked (Pearl 2000; Hitchcock in ch. 3 of this volume; Hausman 1998) a similar strategy is no longer possible when a fundamental theory is applied to the whole universe at once. Now

[24] See Woodward (2003, pp. 127 ff) for additional discussion.

there is no longer anything outside the system being modeled to serve as possible source of interventions and it may be quite unclear how one may legitimately model interventions as part of the system being studied.

As an illustration, consider the claim that (U) the state S_t of the entire universe at time t causes the state S_{t+d} of the entire universe at time $t + d$. On an interventionist construal, this claim would be unpacked as a claim to the effect that under some possible intervention that changes S_t, there would be an associated change in S_{t+d}. The obvious worry is that it is unclear what would be involved in such an intervention and unclear how to assess what would happen if it were to occur, given the stipulation that S_t is a specification of the entire state of the universe. Although I don't claim that it is *obvious* that the relevant interventionist counterfactuals make no sense or lack determinate truth values, it seems uncontroversial that a substantial amount of work would have to be done to explain what these counterfactuals mean.

Commenting on this point, Pearl writes: 'If you wish to include the whole universe in the model, causality disappears because interventions disappear—the manipulator and the manipulated lose their distinction.' (2000, p. 350). While I am less confident than Pearl that causality 'disappears' in these circumstances, I think that it is very plausible that causal ascription becomes less natural and straightforward—increasingly strained—when candidate causes expand to include the state of the entire universe.

4.7 Arrow Breaking

There are several other features of the systems that are studied in the upper-level sciences that make the application of the notion of an intervention seem particularly apt. As we noted in connection with the *ABS* (pressure, barometer, storm) system, a natural way to investigate the causal structure of complex systems is to take them apart. By breaking or disrupting certain causal relationships in a system one may create circumstances in which other causal relationships, if real, will reveal themselves in associations. Thus, if we disrupt the relationship between *A* and *B* by manipulating *B*, we expect any causal relationship between *B* and *S* to show itself in a correlation between *B* and *S* that persists under this manipulation. Similarly, in the drug experiment, the effect of randomization is to replace a situation in which, for example, subjects decide on their own whether or not to take a drug with a situation in which who does or doesn't get the drug is controlled by the randomization process. It seems unproblematic to suppose that the causal influence of the subject's decisions on whether they take the drug is 'turned off' when the randomization is instituted and this assumption is crucial to the inference we draw from the randomized trial.

These assumptions about the possibility of turning off or breaking certain causal influences in order to isolate and investigate others go hand in hand with the fact that causal generalizations on which common sense and the special sciences focus have only limited ranges of invariance and stability. Because these generalizations hold only for a certain range of conditions and break down outside of these, it is possible, either by actively creating situations in which these generalizations break down or finding naturally occurring situations in which this happens, to turn off the causal influences they describe and use this operation to investigate other generalizations or causal relationships that remain intact. Thus we can readily create situations which disrupt the causal connection between atmospheric pressure and the barometer reading or between the experimental subject's own decisions and whether he takes a drug and then determine whether the relationship between barometer reading and the occurrence of the storm, or between drug ingestion and recovery remain intact under this operation. Again, however, it seems less clear how to carry over this idea of breaking some causal influences in order to investigate others into all of the contexts in which theories of fundamental physics apply. As noted above, we think that it is a mark of fundamental laws that they either do not break down at all or break down only in very special and unusual situations—neither experimenters nor nature can create such situations in anything like the range of circumstances in which typical macroscopic causal relationships can be disrupted.[25] Although, as Woodward (2003, s. 3.5) argues, we can sometimes appeal to our theories themselves to tell us what would happen under interventions that are counter-nomic, the fact remains that in many physics contexts there may be no physically realistic operation corresponding to placing some variable of interest entirely under the control

[25] This is not to say that there is nothing that looks like arrow-breaking in experimentation in physics—for example, one may shield an apparatus from electromagnetic forces that would otherwise be operative.

of an intervention variable, and breaking all other causal arrows directed into it.

There is yet another feature of the notion of an intervention that influences its application to the sorts of systems that are studied in the special sciences. Consider again the use of a randomized experiment to test the claim that a drug produces recovery from a certain disease. I said above that (in the absence of improbable coincidences) such an experiment will approximate the conditions for an intervention on treatment with the drug with respect to recovery. Recall that one condition for a successful intervention is that the intervention I on X with respect to Y should not cause Y via a route that does not go through X, and that I should be independent of any variable Z that causes Y but not via a route that goes through I and X. However, there is an apparently natural line of thought, echoing the argument in Section 5, that questions whether these conditions are ever likely to be satisfied—either in the randomized experiment under consideration or any other realistic case. Consider the microstates s_i, characterized in terms of quantities provided by fundamental physics, that realize the values of the variable 'recovery/non-recovery' for each subject *i*. Surely, it might be argued, these states s_i will themselves be causally influenced (via a route that does not go through the putative cause variable, ingestion (or not) of the drug) by the states of the micro-variables that realize the values of the intervention variable and by the microstates of causes of the intervention variable. For example, each occurrence of the coin flip that implements the randomization will alter the position of various elementary particles and will have the consequence that various forces on the variables s_i will be different to what they otherwise would be. For all we know, this fact will show up in some correlation between these sets of variables on repeated flips. Moreover, some of these forces will operate independently of whether the drug does or does not cause recovery. A similar point will hold for whatever micro-variables influence the outcome of the coin flip. In short (it might be argued), at the level of fundamental physics, events will be causally interconnected in a way that precludes the satisfaction of the conditions for an intervention.[26]

[26] Michael Friedman, among others, has suggested in conversation that something like this is true. Another (perhaps better) way of putting the worry is that the contrast between those variables that influence the effect only through the cause and those that instead influence the effect via a route that does not go through the cause—a contrast that is at the heart of the notion of an intervention—becomes

I believe that this objection will only seem plausible if we fail to take seriously the observations about coarse-graining made earlier. The claims in the previous paragraph about the independent causal influence of the micro-variables realizing the intervention on the micro-variables realizing recovery are perfectly correct but it does *not* follow that the intervention itself or the causes of it have a causal influence on recovery that is independent of treatment with the drug. If $s_1 \ldots s_n$ are micro-variables that realize the macroscopic, coarse-grained variable X and $s_1^* \ldots s_n^*$ are micro-variables realizing the macroscopic coarse-grained variable Y, it is perfectly possible for some instantiations of some of the s_i to causally influence and be correlated with some instantiations of some of the s_i^* and yet for it to be false that X causally influences or is correlated with Y. Suppose that we have a population of patients with a disease, some of whom will recover and some of whom will not. For each patient we flip a coin and record the results but these results are not used to determine what treatment, if any, the patients will receive. Instead we simply observe whether or not each patient recovers. For just the reasons described in the previous paragraph, some of the micro-level variables the values of which realize the coin flip will causally influence the micro-level variables realizing instances of the recovery variable but of course it does not follow (and we do not expect that) there will be a correlation between the outcome of the coin flip for individual patients and whether they recover—we don't think one can use the outcome of the coin flip to predict who will recover. In just the same way we don't think that we can use the outcome of coin flips to predict the future of the stock market or who will win the next US presidential race despite the presence of causal influences among the micro-realizations of these variables. As already noted, it is possible for causal and statistical independence among groups of coarse grained variables to emerge from a web of complicated causal dependencies in which everything is influenced by everything else (in its backward light

unclear at the level of fundamental physics. That is, the whole notion that one variable might affect another via multiple distinct routes is itself a consequence of our adoption of a coarse-grained perspective and the distinctness of different routes itself disappears at a fine-grained level, where the correct causal representation (if there is one) is just a chain structure in which a succession of single arrows connects one total state of the universe to another. This last representation eliminates the worry about independent influences on the effect that do not go through the cause, but has the result that everything in the absolute past of an event is now causally relevant to it. Thanks to Chris Hitchcock for helpful conversation on this point.

cone and to some degree or other) at a more fine-grained level. This in turn makes it possible to actually carry out interventions, which require such independence.

4.8 Distinguishing Causes and Correlations

I noted in Section 4.1 that in upper-level sciences in which causal talk plays a major role we often face is a generic inference problem that looks like this: We know that two variables *X* and *Y* are correlated but we don't yet know what their causal relationship is—whether *X* causes *Y*, *Y* causes *X*, whether there is a third variable or set of variables which are common causes of both *X* and *Y*, whether the correlation is present only in a sample that is not representative of the larger population from which it has been taken, and so on. To an important extent, the role played by our notion of causation is tied up with elucidating the differences among these possibilities. It is interesting that the same inference problem seems to arise much less often in fundamental physics. While one can think of possible exceptions (e.g. perhaps the status of thermodynamic asymmetries or the status of various boundary conditions on the whole universe imposed in cosmology), physicists are not usually in the epistemic position of knowing that, say, a regularity holds globally among fundamental physical quantities but not knowing whether this regularity represents a causal or nomological relationship or whether the regularity is a mere correlation is produced by some third variable. Instead, the more usual situation is this: although it may not be clear whether some proposed generalization holds globally, there is little doubt about whether if it is true, it will fall into the category of a 'law' or (less commonly) that of an 'accidental' generalization. Consider the equations specifying the coupling of gravity to matter associated with the various gravitational theories (e.g. the Brans-Dicke theory) that once were regarded as alternatives to General Relativity. No one appears to have doubted that if one of these equations had turned out to be true, it would have been a law of nature—that is, no one took seriously the possibility that the equation might be true but only accidentally so, in the manner in which the relationship between parental income and child's scholastic achievement might be merely accidentally true or a mere correlation. So the inference problem that looms so large in the special sciences and which helps to give causal talk its point seems somehow much less pressing in fundamental physics.

What accounts for this difference? One factor that seems to be at work in fundamental physics contexts is the availability of a great deal of detailed background knowledge/expectations that guides our decisions about whether a generalization is appropriately regarded as a candidate for a law or merely accidentally true. This includes information to the effect that laws should satisfy various symmetry requirements—if a generalization fails to satisfy these requirements it is unlikely to be a law, regardless of whatever other descriptive virtues it may possess. It may also include information to the effect that the generalization of interest is derivable from known laws and initial conditions that are identified by theory as holding pervasively but merely contingently as may be the case for the second law of thermodynamics and certain cosmological regularities.

The situation in the special sciences is quite different. First, causal generalizations as well as generalizations that describe mere correlations do not even purport to hold globally. True causal generalizations in the special sciences are typically restricted to various more or less specialized systems and break down under a variety of conditions, not all of which are well understood. In the special sciences, the causal/accidental distinction is a distinction *within* the category of non-universal generalizations with many exceptions. Second, in many contexts in the special sciences, we lack the kind of background theoretical knowledge that would provide a basis for sorting generalizations into the categories of causal vs. merely accidental or correlational. At least in part for these reasons, the epistemic problem of deciding into which of these categories a candidate generalization falls has a kind of salience in the special sciences that it does not have in fundamental physics. With this comes a greater use of concepts and patterns of reasoning designed to mark this distinction.

4.9 Are Macroscopic Causal Relationships Real?

My strategy so far has been to draw attention to features of the systems typically studied in the special sciences that make characterizations that appeal to causal notions particularly useful and illuminating. I'm fully aware, however, that this strategy will seem unsatisfying to the metaphysically minded.

Even if it is granted that causal description is sometimes useful, this leaves untouched (it will be said) the question of whether causal relationships are 'real' or 'objective'. What if anything do the arguments in Sections 4.1–4.8 suggest about the literal truth of causal claims about macroscopic systems? Should such claims be construed purely instrumentally, as nothing but helpful fictions?

Let me approach this issue by means of an analogy. Consider the role of chance in systems that are governed by deterministic laws. In particular, consider a gambling device like a roulette wheel, the operation of which (let us suppose) is governed at the relevant level of analysis by deterministic laws and yet which generates outcomes that are (or look as though they are) independent and occur with stable probabilities strictly between zero and 1. How is this possible?

The broad outlines of an answer to this question go back at least to Poincaré and have been set out (and generalized) by Michael Strevens in a recent interesting book (2003). Loosely described, Strevens' treatment appeals to the following ideas. First, consider the 'evolution functions' that map the initial conditions in the system of interest that are relevant to the final outcome onto the various particular values of the macroscopic outcome variable. In the case of the roulette wheel, the initial conditions will include, for example, the initial position of the wheel and the angular momentum imparted to it by the croupier, and the outcomes will be a particular number or color. Assume that the dynamics of the system are such that it exhibits 'sensitive dependence to initial conditions'—more specifically, assume that nearby regions in the phase space of initial conditions map onto different outcomes, and that for an appropriately chosen partition of the phase space into small contiguous regions, the volume of the regions that are mapped into each of the outcomes is constant, or approximately so, within each cell of the partition. Then in repeated trials that impose any distribution of initial conditions that is appropriately 'smooth' (as Strevens calls it—this means that the probability density of the initial conditions is approximately constant within each cell of the partition), one will get outcomes that are independent and have stable probabilities. But why the restriction to smooth distributions? Although Strevens does not quite put it this way, one natural motivation is this: It is plausible to assume that in the case of real life gambling devices like roulette wheels any macroscopic (hence spatially and temporally extended) system (like the croupier) can only intervene on the wheel in a coarse-grained or macroscopic manner: that is, it cannot impose a non-smooth distribution on the relevant initial conditions of the system. It cannot, for example, fix (via a sort of Dirac delta function operation) a particular point value for the initial conditions of the wheel with probability 1, so that all other initial conditions receive zero probability.

Assume for the sake of argument that this claim is correct. What follows about the objective reality of chances (that is, the reality of non-trivial chances strictly between zero and 1) for such devices? If our criterion for the objective reality of such chances is whether they appear in fundamental physical laws then the answer is clear: the chances we associate with such devices are not objectively real, however useful they may be for summarizing the behavior of the devices and guiding betting behavior. Because the underlying laws are deterministic, the objective chance of, say, 'red' on any given spin of the wheel is always either zero or 1. Any other value for the probability of this outcome must be understood as a 'subjective' or 'merely epistemic' probability, reflecting our ignorance of exact initial conditions.

Looked at another way, however, this assessment is misleading. The non-trivial chances we ascribe to outcomes are a reflection of an interaction between perfectly objective facts about the dynamics of gambling devices and other facts (arguably also equally objective) about the kinds of distributions of initial conditions that a macroscopic agent or process is able to impose. Moreover, given these facts, it may well be true that there is nothing that any macroscopic process (the hand of any croupier or any similar sized intervention) can do in the way of imparting a spin to the wheel which will produce deviations from the ascribed chances, as well as no observations or measurements that might be made by a macroscopic observer that will allow for the prediction of anything more about outcomes than is already recorded by the chances. So if we mean by objective probabilities, probabilities that reflect patterns that will obtain under any distribution over initial conditions that a macroscopic agent is able to impose or learn about—that is, probabilities that are invariant/stable under changes in *these* distributions, even if they are not invariant/stable under other, conceivable distributions—then there are non-trivial objective probabilities associated with the roulette wheel. It would be quite mistaken to suppose that these probabilities are 'subjective' in the sense that, say, they are matters of individual taste, limited only by considerations of coherence, or that, given the facts about the dynamics and the initial conditions it is possible to impose, there is no fact of the matter about which probability assignments are correct. Instead the probabilities are 'real' in the straightforward sense that they reflect constraints not just on what macroscopic agents are able to learn about but also what they are able to do.[27]

My suggestion is that at least in some respects the status of causal relationships in macroscopic systems is similar. If our criterion for 'objectively real' is 'found (or grounded in a direct way) in the laws of fundamental physics alone', then, as we have seen, it is dubious that most macroscopic causal relationships are 'real'. Or, to put the point more cautiously, it seems doubtful that we will find, for each true causal claim in common sense and the special sciences, counterpart relations in fundamental physical laws, with all of the features that we ascribe to macroscopic causes. This is because macroscopic causal relationships do not depend just on facts about fundamental laws but also reflect other considerations as well—for example, the coarse graining operations associated with our status as macroscopic agents and the frequency with which various initial conditions happen to occur in our spatiotemporal vicinity. Among other things, coarse-grained factors and events, with ill-defined boundaries and spatiotemporal relations, are not plausibly regarded as instances of fundamental laws.

Just as with the chances we ascribe to deterministic gambling devices, nothing prevents us from adopting the grounded-in-physical-laws-alone criterion for what is 'real' as a stipulative definition. However, we need to take care that we do not read more into this stipulation than is warranted. In particular, as nearly as I can see, it is consistent with the 'unreality' of macroscopic causation, in the sense associated with the criterion just

[27] The anonymous referee worries that the roulette wheel exhibits merely a 'counterfeit' notion of non-trivial objective chance. But contrast the claim that the probability of red equals 0.5 on the next spin of a roulette wheel satisfying the conditions described above with the claim that, say, the probability of war between India and Pakistan in the next ten years is 0.5. In the latter case, we have no corresponding story about what would constitute repeated trials with the same chance set-up, and no story about how the interaction between an underlying dynamics and the initial conditions that a macroscopic agent (or other process) is able to impose yields stable frequencies etc. There are real differences between the processes that underlie the generation of outcomes in the case of the roulette wheel and the processes that will generate war (or not) between India and Pakistan. It is these differences that make us think that ascription of chances in the latter case is far more 'subjective' than in the former case. So while one can certainly stipulate a meaning for 'objective chance' according to which non-trivial objective chances require indeterminism, this has the unfortunate consequence of collapsing important distinctions among the behavior of deterministic systems.

described, that there are, in a straightforward sense, facts about whether manipulating macroscopic variable *X* will be associated with changes in macroscopic variable *Y* or whether instead the observed association between *X* and *Y* is a mere correlation. Similarly, there are facts about whether a relationship between *X* and *Y* that is exploitable for purposes of manipulation would continue to hold across various sorts of changes in background conditions. Moreover, such facts can be discovered by ordinary empirical investigation. If the line taken in this paper is correct, macroscopic causal claims (like 'chances' in a deterministic world) reflect complicated truths about (i) an underlying microphysical reality and (ii) the relationship of macroscopic agents and objects to this world. This second ingredient (ii) gives macroscopic causal talk a number of its characteristic features—coarse-grainedness, a focus on small worlds where this is a possibility of outside intervention, reliance on contingent facts about initial and boundary conditions and so on—but it does not make such claims 'subjective' in the sense of not being controlled by evidence, dependent on the idiosyncratic tastes or interests of individual investigators and so on. Just as the notion of chance seems pragmatically unavoidable when dealing with systems like roulette wheels (it is not as though we have some alternative available which works better), so also for 'causation' when dealing with many macroscopic systems.

4.10 Conclusion

I conclude with a puzzle/concession and a restatement of a moral. The former has to do with the role of explanation in physics. Suppose that causation (at least when construed along the interventionist lines that have been described) does not play a foundational role in fundamental physics. What follows for how we should think about explanation in such contexts? Are all explanations causal with the apparent result that at least in some respects and contexts, fundamental physics does not provide explanations? Does physics instead provide non-causal explanations and if so, how should we understand the structure of these?[28] I am not sure what to think about these questions, but also do not think that this uncertainty is in itself a

[28] I am grateful to the anonymous referee for raising these questions, which are undeniably relevant and important.

good reason to reject the conclusions about the role of causation in physics reached above.

Next, the moral. This has been implicit in earlier portions of my discussion, but merits explicit underscoring. It is a popular idea that true causal (and counterfactual claims) from everyday life and the special sciences cannot be 'barely true' but instead require grounding (or 'truth-makers') in fundamental physical laws. This idea strikes me as arguably correct if it is interpreted in the following way: given a true garden variety causal claim, there will be some associated in—principle physical explanation (or story or account, to use more neutral words) for its holding, and this will include, among other factors, appeal to fundamental laws. However, the idea in question often seems to be understood in a different and more restrictive way: as the claim that reference to fundamental laws alone (together with the claim that the events or factors related in the causal claim 'instantiate' the law or bear some other appropriate relationship to it) gives us all that is needed to state the grounds or truth-makers for causal claims. Here the idea is that the causal part of the content of all causal claims is somehow grounded in the fundamental laws themselves with nothing else required. If the argument of this paper is correct, this second interpretation of the grounding idea is mistaken. Typically, the grounds or truth-makers for upper-level causal claims like '*Cs* cause *Es*' or 'particular event *c* caused particular event *e*' will involve many additional factors besides laws (and besides facts about whether *C*, *E*, *c* and *e* instantiate laws or are part of conditions that instantiate laws etc.). These additional factors will include very diffuse, messy, and non-local facts about initial and boundary conditions that do not obtain just as a matter of law and have little to do with whatever underlies or realizes *C*, *E*, *c* or *e* themselves (recall the discussion in s. 4.4). If so, attempts to provide sufficient conditions for *Cs* cause *Es* along the lines of '*Cs* are (or are part of conditions that are) linked by fundamental laws to *Es*' are unlikely to be successful.

References

Armstrong, D. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.

Arntzenius, F. (1990). 'Common causes and Physics', *Synthese*, 82: 77–96.

Cartwright (1979). 'Causal Laws and Effective Strategies', *Noûs*, 13: 419–37.

Chu, T., Glymour, C., Scheines, R. and Spirtes, P (2003). 'A Statistical Problem for Inference to Regulatory Structure from Associations of Gene Expression Measurement with Microarrays', *Bioinformatics 19*, pp. 1147–52.

Coleman, J. and Hoffer, H. (1987). *Public and Private High Schools*. New York: Basic Books.

Davidson, D. (1967). 'Causal Relations'. *Journal of Philosophy*, 64: 691–703.

Dowe, P. (2000). *Physical Causation*. Cambridge, UK: Cambridge University Press.

Field, H. (2003). 'Causation in a Physical World', in M. Loux and D. Zimmerman (eds). *Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.

Frisch, M. (2000). '(Dis-)Solving the Puzzle of the Arrow of Radiation', *British Journal for the Philosophy of Science,* 51: 381–410.

Frisch, M. (2002). 'Non-Locality in Classical Electrodynamics', *British Journal for the Philosophy of Science*, 53: 1–19.

Glymour (forthcoming). 'Mental Causation and Supervenient Science'.

Gopnik, A. and Schulz, L. (forthcoming). *Causal Learning: Psychology, Philosophy and Computation*. Oxford: Oxford University Press.

Granger, C. W. J. (1998). 'Granger Causality', in J. Davis, D. Hands, U. Maki (eds), *Handbook of Economic Methodology*. Aldershot: Edward Elgar, 214–16.

Haavelmo, (1944). 'The Probability Approach in Econometrics', *Econometrica*, 12 (Supplement): 1–118.

Hausman, D. (1998) *Causal Asymmetries*. Cambridge, Cambridge University Press.

Hausman, D. and Woodward, J. (1999). 'Independence, Invariance and the Causal Markov Condition', *The British Journal for the Philosophy of Science,* 50: 521–83.

Hitchcock, C. (this volume). 'What Russell got Right'.

Hoover, K. (1988). *The New Classical Macroeconomics*. Oxford: Basil Blackwell.

Jackson, J. D. (1999). *Classical Electrodynamics*. New York: Wiley.

Menzies, P. and Price, H. (1993). 'Causation as a Secondary Quality'. *British Journal for the Philosophy of Science*, 44: 187–203.

Norton, J. (this volume). 'Causation as Folk Science'.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge, Cambridge University Press.

Price, H. (1991). 'Agency and Probabilistic Causality', *British Journal for the Philosophy of Science*, 42: 157–76.

Rueger, A. (1998). 'Local Theories of Causation and the Aposteriori Identification of the Causal Relation', *Erkenntnis*, 48: 25–38.

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

Schaffer, J. (2000). 'Causation by Disconnection', *Philosophy of Science*, 67: 285–300.

Sklar, L. (1993). *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Smith, S. (forthcoming). 'Armstrong on the Relationship between Causation and Laws'.

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd edn. New York, NY: MIT Press.

Strevens, M. *Bigger Than Chaos*. (2003). Cambridge: Harvard University Press.

Strotz, R. H. and Wold, H. O. S. (1960). 'Recursive vs. Non-recursive Systems: An Attempt at a Synthesis', *Econometrica*, 28: 417–27.

Tomasello, M. & Call, J. (1997). *Primate Cognition*. New York: Oxford University Press.

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

Woodward, J. (forthcoming a) 'Sensitive and Insensitive Causation', *Philosophical Review* 115 (1).

Woodward, J. (forthcoming b) 'Interventionist Theories of Causation in Psychological Perspective', in A. Gopnik and L. Schulz (eds), *Causal Learning: Psychology, Philosophy and Computation*. New York: Oxford University Press.