Salmon on Explanatory Relevance
Author(s): Christopher Read Hitchcock
Source: *Philosophy of Science*, Vol. 62, No. 2 (Jun., 1995), pp. 304-320
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/188436
Accessed: 08/04/2014 13:05

DISCUSSION:

SALMON ON EXPLANATORY RELEVANCE*

CHRISTOPHER READ HITCHCOCK†‡

Department of Philosophy
Rice University

One of the motivations for Salmon's (1984) causal theory of explanation was the explanatory *irrelevance* exhibited by many arguments conforming to Hempel's covering-law models of explanation. However, the nexus of causal processes and interactions characterized by Salmon is not rich enough to supply the necessary conception of explanatory relevance. Salmon's (1994) revised theory, which is briefly criticized on independent grounds, fares no better. There is some possibility that the two-tiered structure of explanation described by Salmon (1984) may be pressed into service, but more work would have to be done. Ironically, Salmon's difficulties are similar to those suffered by his seventeenth-century predecessors.

1. Introduction. A generation ago, Hempel's covering-law models of scientific explanation were widely accepted among philosophers of science. One of the critics who has been most active in dissolving this consensus is Wesley Salmon. Many of Salmon's criticisms have centered on a common theme: Hempel's models fail to capture the intuitive relation of *explanatory relevance* that holds between that which explains and that which is explained. Salmon has argued that this relation is irreducibly *causal*, and he has built his own account of scientific explanation around a well-developed theory of causality (Salmon 1984), which he has recently revised (Salmon 1994). I will argue that Salmon's causal theory of explanation—in both its original and revised form—also fails to capture the intuitive relation of explanatory relevance.

2. D-N Explanation and its Discontents. The modern history of the philosophical study of scientific explanation began with Hempel and Oppenheim's "Studies in the Logic of Explanation" (Hempel and Oppenheim 1948). In this essay, Hempel and Oppenheim put forward what Hempel would later call the *Deductive-Nomological* (*D-N*) model of explanation. In his later essay, "Aspects of Scientific Explanation,"

---

*Received July 1994; revised September 1994.

†I am grateful to Wes Salmon and Jim Woodward for their many discussions with me on this topic.

‡Send reprint requests to the author, Department of Philosophy, Rice University, MS 14, Houston, TX 77005-1892, USA.

Philosophy of Science, 62 (1995) pp. 304–320

Hempel (1965) defended the original theory, and augmented it with models of statistical explanation. He summarized his conception of explanation as follows:

> . . . all scientific explanation involves, explicitly or by implication, a subsumption of its subject matter under general regularities; . . . it seeks to provide a systematic understanding of empirical phenomena by showing that they fit into a nomic nexus. (Hempel 1965, 488)

Let E be a statement that describes some event that is to be explained—the *explanandum*. A *D-N* explanation of this event has the following form:

$$\begin{array}{l} C_1, C_2, \ldots, C_k \\ L_1, L_2, \ldots, L_r \\ \therefore E. \end{array}$$

$C_1, C_2, \ldots, C_k$ are statements describing particular states of affairs, which may be thought of as initial conditions (although they are not officially required to temporally precede E), and $L_1, L_2, \ldots, L_r$ are general laws; together, these form the *explanans*. Thus an event is explained by showing that its occurence can be deduced from some set of initial conditions together with general laws. The account is called 'deductive-nomological' because it involves deduction from laws. Hempel considers the *D-N* model (as well as his statistical models) to embody what Dray (1957) has called a "covering-law" conception of scientific explanation: the explanandum is explained by being subsumed under general laws.

Salmon's writings (e.g., 1965, 1971, 1984) contain a sustained attack on the *D-N* model and the other covering-law models of explanation. This attack exploits several examples, now well known, where the conditions of *D-N* explanation seem to be met, but where the purported explanation seems to be defective. Here are two of the more famous examples:

> This sample of table salt dissolves in water, for it has had a dissolving spell cast on it, and all samples of table salt that have had dissolving spells cast on them dissolve in water. (Kyburg 1965, 147)
>
> John Jones avoided becoming pregnant during the past year, for he has taken his wife's birth control pills regularly, and every man who regularly takes birth control pills avoids pregnancy. (Salmon 1971, 34)

In each case, the condition described in the explanans—the hexing of the salt and the faithful consumption of contraceptives—is *irrelevant* to the outcome in question.

Salmon (1971) attempted to analyze this notion of explanatory relevance in terms of *statistical* relevance. Among men, the consumption of oral contraceptives is statistically irrelevant to pregnancy: those men who take birth control pills become pregnant with exactly the same statistical frequency as those who do not. Analogously, among samples of salt, hexing is statistically irrelevant to dissolution. By 1984, however, Salmon had concluded that statistical relevance relations alone could not provide a satisfactory account of explanation: explanatory relevance is an irreducibly causal concept. Salmon summarizes his conclusion by advocating the replacement of two words in the summary of Hempel's account cited above:

> . . . my suggestion for modification would be to substitute the words "*how* they fit into a *causal* nexus" for "*that* they fit into a *nomic* nexus." (Salmon 1984, 19)

Salmon then aims to provide an account of the causal nexus: it is a network of *causal processes* and *causal interactions*.

3. Causal Processes and Causal Interactions. Salmon's theory of causation is influenced, in part, by foundational work in the special theory of relativity. Presentations of that theory sometimes commence with the following axiom: nothing can travel faster than light. In order for this axiom to be true, the scope of the quantifier 'nothing' needs to be restricted, and therein lies the connection with causation. By pointing a laser at a sufficiently distant wall, and rotating the laser with a sufficient (but subluminal) velocity, it would be possible to make the spot of light cast on the wall move faster than the speed of light. But this moving spot of light would not instantiate the sort of supraluminal velocity proscribed by the special theory of relativity. In particular, this moving image could not be exploited to send messages from one point on the wall to another. For example, suppose that a filter is held in front of a small section of the wall, thus diminishing the intensity of the light striking that spot. The moving light spot will not retain its diminished intensity as it sweeps out its path, so this 'message' will not be transmitted. The anthropocentric reference to *our* inability to transmit messages may be avoided; what is essential is that the image be incapable of conveying any sort of causal influence. The notion of a *causal process*—a process capable of carrying causal influence—is therefore of extreme physical importance.

Salmon's theory takes the notion of a process as a primitive. Informally, a process is a path through space-time that exhibits a certain continuity in its properties. A moving shadow is a process: it may be slightly darker in some regions, or grow lighter with time, but if there is a sudden difference in brightness between two nearby points in space-time, then one of those points lies outside the spatiotemporal boundaries of the moving shadow. Processes come in two varieties: pseudoprocesses—such as the moving shadow or the moving spot of light on the wall—and causal processes. A causal process, according to Salmon (1984), is one that is capable of transmitting a mark, where mark-transmission is defined as follows:

> MT: Let P be a process that, in the absence of interactions with other processes, would remain uniform with respect to a characteristic Q, which it would manifest consistently over an interval that includes both of the space-time points A and B ($A \neq B$). Then, a mark (consisting of a modification of Q into Q'), which has been introduced into process P by means of a single local interaction at point A, *is transmitted* to point B if P manifests the modification Q' at B and at all stages of the process between A and B without additional interventions. (Salmon 1984, 148)

This definition of mark-transmission is counterfactual in nature: whether a process transmits a mark depends upon how it *would* behave if unmarked. Similarly, the definition of a causal process also has a counterfactual element: a causal process is one that is *capable* of transmitting some type of mark regardless of whether any marks are *in fact* being transmitted. Salmon argues that the required counterfactuals are subject to empirical test, and thus unproblematic. We know, for example, that a baseball is a causal process; for when baseballs are scuffed by small pieces of sandpaper—as occasionally happens when pitchers use illegal means to achieve baffling aerodynamic effects—these scuffs remain on the ball.

This account of causal processes provides a theory of causal *propagation*, the transmission of causal influence. Salmon is also interested in causal *production*. According to Salmon, many changes in causal processes originate in causal interactions, which sometimes occur when causal processes intersect. The definition of a causal interaction presented in Salmon (1984) is as follows:

> CI: Let P_1 and P_2 be two processes that intersect with one another at the space-time point S, which belongs to the histories of both. Let Q be a characteristic that process P_1 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P_1) if the intersection with P_2 did not occur; let R be a characteristic that process P_2 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P_2) if the intersection with P_1 did not occur. Then, the intersection of P_1 and P_2 at S constitutes a causal interaction if:
>
> 1) P_1 exhibits the characteristic Q before S, but it exhibits a modified characteristic Q' throughout an interval immediately following S; and

> 2) P_2 exhibits the characteristic R before S, but it exhibits a modified characteristic R' throughout an interval immediately following S. (p. 171)
>
> . . . In order to transform *CI* into a condition that is necessary as well as sufficient, we need simply to say that a causal interaction occurs if and only if there exist characteristics Q and R that fulfill the conditions stated previously. (p. 174)

An intersection of causal processes is not always a causal interaction. If two light beams cross, there is a local distortion of each one, but each beam continues on as if there had been no intersection. By contrast, a collision between two cars is a causal interaction, each car exhibiting a modified characteristic which persists after the collision: bent fenders. Note that condition *CI*, like *MT*, involves a counterfactual component: changes in the properties of intersecting processes constitute a causal interaction only if those changes would not have occurred without the intersection. While Salmon presents the two definitions in this order for heuristic reasons, *CI* is logically prior to *MT*: introduction of a mark into a causal process is a special case of causal interaction (Salmon 1994, 298).

A causal nexus, then, is a physical network consisting of causal processes and interactions. What is it to show how an explanandum fits into a causal nexus? Presumably it is to describe the location of that explanandum within such a nexus—to provide a kind of causal map. Salmon expands upon this answer by distinguishing between two different aspects of explanation:

> Suppose we want to explain some event *E*. We may look at *E* as occupying a finite volume of four-dimensional space-time [see figure 1]. If we want to show why *E* occurred, we fill in the causally relevant processes and interactions that occupy the past light cone of *E*. This is the etiological aspect of our explanation; it exhibits *E* as embedded in its causal nexus. If we want to show why *E* manifests certain characteristics, we place inside the volume occupied by *E* the internal causal mechanisms that account for *E*'s nature. This is the constitutive aspect of our explanation; it lays bare the causal structure of *E*. (Salmon 1984, 275)

Salmon (1994) has recently abandoned definitions *MT* and *CI* in favor of a new characterization of causal processes and interactions. Because it is not as familiar as the old theory, I will postpone discussion of the new account until Section 5. So far as I can tell, the new characterization of the causal nexus has not affected Salmon's view of the role that the causal nexus plays in explanation.

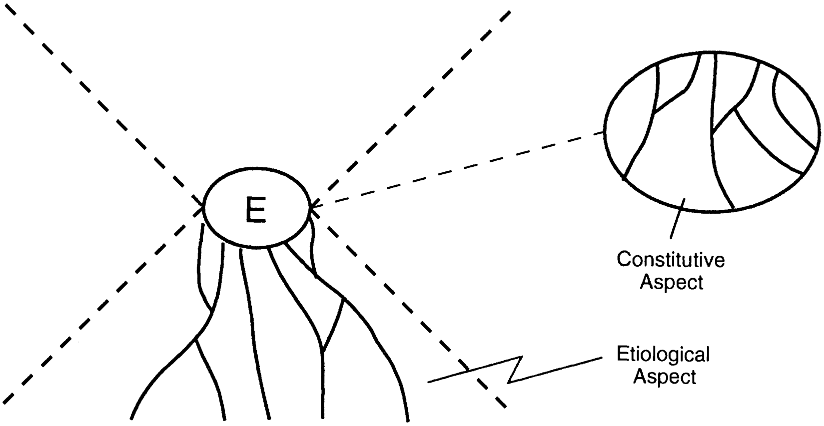

Fig. 1. The two aspects of explanation.

A caveat: several authors have suggested that there are fields of inquiry where explanations do not accord with Salmon's account: formal linguistics, pure mathematics, general relativity, and quantum mechanics to name a few (see Kitcher 1989 and Woodward 1989 for examples). These tend to be areas where the notion of causal explanation may not be appropriate. Yet, if Salmon succeeds only in characterizing a subclass of scientific explanations—*causal* explanations—he will still have accomplished a great deal. In the sequel, I will restrict attention to examples in which causal explanation is clearly appropriate.

4. Relevance Revisited. One of Salmon's motives for abandoning Hempel's covering-law models of explanation is their failure to capture the notion of explanatory *relevance*. Does Salmon's theory do any better? Let us look again at the two counterexamples to Hempel's theory. *If* in the process of hexing the salt, the (person dressed up as a) witch never initiated a causal process that interacted with the salt, then it would seem that Salmon's theory would give us grounds for rejecting the hex as explanatorily irrelevant. Let us suppose, however, that during the casting of her spell, our would-be sorceress touched the salt with a wand. This is a genuine causal interaction, so what would justify our rejection of it as irrelevant? Similarly, there were a series of causal interactions between John Jones and some birth control pills: what entitles us to reject these as explanatorily irrelevant? Salmon's answer would appear to be "nothing," but let us explore the question a little more deeply.

One difficulty with these two counterexamples to Hempel's theory is that the causal processes and interactions involved in each are quite complex, making it difficult to see how Salmon's account might treat them. The following example is a more sanitized one: In a game of pool, the cue ball, eight ball, and corner pocket all lie on a straight line, and a player (let us call her "Pennsylvania Slims") manages to hit the cue ball in the right direction so as to sink the eight ball into the corner pocket. Before hitting the cue ball, Ms. Slims chalked her cue stick with blue chalk. Some of this chalk came off on the cue ball when it was struck. We may now construct a *D-N* 'explanation' of the sinking of the eight ball, citing as initial conditions the linear momentum imparted to the cue ball *and* the blue spot that it acquired, together with the law that physical systems containing objects with blue spots have their linear momentum conserved. (It may be objected that this is not a law, but a similar objection may be raised against the original counterexamples.) This defective explanation involves a failure of relevance of the familiar sort: the blue spot was irrelevant to the final position of the eight ball.

How would Salmon's theory handle this example? Looking into the past light cone of the sunken eight ball we find: a causal interaction between the cue stick and the cue ball; the moving cue ball, which is a causal process; a causal interaction between the cue ball and the eight ball; and another causal process consisting of the moving eight ball. The explanation will then exhibit the location of the explanandum within this network of causal processes and interactions. (Note that we have ignored the constitutive aspect of explanation here, but the microstructure of the eight ball in the corner pocket clearly does not show why the eight ball manifests the characteristic of being in the corner pocket.) Where, in any of this, are we to find the desired relations of explanatory relevance?

The intuitive relation of explanatory relevance does not hold between regions of space-time: it holds between the *properties* instantiated in certain regions of space-time (or perhaps between the *propositions* that certain properties are instantiated in certain regions of space-time). We judge that the *linear momentum* of the cue ball is relevant to the final *location* of the eight ball, but that the blue color on (part of) the cue ball is not. Such properties of causal processes do have a role to play in Salmon's theory. A process is causal if it is capable of transmitting a mark, where a mark is some change in one of the properties of a process, and a mark is transmitted if the new property is manifested for some duration of time. Causal interaction is likewise characterized in terms of modifications of properties of processes. Nonetheless, Salmon's explanations do not seem to *cite* these properties. In locating the sunken eight ball within its causal nexus, the only property that is ascribed to the cue ball is that of being a causal process.

Reflecting on the counterexamples to the *D-N* model, it is hard to resist the conclusion that our intuitive judgments of explanatory relevance correspond closely to our judgments about the truth of counterfactuals: the salt would have dissolved even if it had not been hexed; John Jones would have avoided pregnancy even if he had not swallowed the pills; the eight ball would have landed in the corner pocket even if the cue ball had not been marked with a blue spot. This is not to say that explanatory relevance ought to be analyzed in terms of counterfactual dependence. Nor is it to say that our judgments of explanatory relevance correspond more closely to our judgments of counterfactual dependence than, say, to our judgments about statistical relevance. Skyrms (1984) argues persuasively that our judgments of statistical relevance and counterfactual dependence (and many other things besides) coincide in ordinary cases. I wish only to draw the rather weak conclusion that a successful account of explanation had better make the relation of explanatory relevance look roughly like that of counterfactual dependence. Salmon's characterization of causal processes and interactions does have the relation of counterfactual dependence—or something suitably close to it—built in. But it is built in in such a way that descriptions of the causal nexus are not specific in their counterfactual commitments. For example, in stating that a process is a causal process rather than a pseudoprocess, it is asserted only that *some* later properties of the process are dependent upon *some* earlier properties; no information is given about what these properties are. Our demand that explanations provide *relevant* information requires something stronger—that we be told *which* earlier properties the properties specified in the explanandum depend upon.

A natural response at this point would be that the explanatorily relevant properties are those in virtue of which processes and interactions in the nexus satisfy conditions *MT* and *CI*. This does not help: when it comes to categorizing a process as causal, for example, any transmitted mark is as good as any other. Consider the moving cue ball, one of the causal processes involved in our central example. What makes this a causal process rather than a pseudoprocess? The moving cue ball is capable of transmitting a mark. The cue ball would not have had a blue spot had it not interacted with the cue stick; thus, this interaction *marked* the cue ball with a blue spot. The cue ball retained the blue spot as it moved, so this mark was *transmitted*. But this was not the only mark transmitted by the cue ball: in striking the cue ball, Ms. Slims gave it a linear momentum it would not otherwise have had, and this change in linear momentum counts as a mark as well. Since the cue ball continued to move with (approximately) the amount of linear momentum that was imparted to it in its interaction with the cue stick, this mark was also transmitted. For purposes of constituting the cue ball as a causal process, both of these marks are on a par. From the perspective of explanation, however, these two marks are not on a par: the linear momentum of the cue ball was *relevant* to the eight ball's landing in the corner pocket; the blue spot on the cue ball was *irrelevant* to this outcome.

As an illustration of how a more satisfactory account of explanatory relevance might proceed, consider the work of James Woodward (1979, 1984). Woodward argues that explanations must provide the resources for answering 'what-if-things-had-been-different' questions. The application of this idea to singular causal explanations is straightforward: if something like Lewis's counterfactual theory of causation is close to being right, then to cite the cause of an event is to provide information about what would have happened had the cause not occurred. But Woodward has also argued that his account provides a diagnosis of why *some* derivations from initial conditions plus laws strike us as explanatory, while other instances fitting the *D-N* schema are defective. Typically, the successful *D-N* explanations involve laws such as Newton's law of gravitation or Coulomb's law that are expressed in *functional* form. In these cases, the explanandum is explained, *not* by subsuming it under laws of nature, but by showing how the outcome in question is sensitive to the nature of the initial conditions cited. He states this "requirement of functional interdependence" as follows:

> The law occurring in the explanans of a scientific explanation of some explanandum *E* must be stated in terms of variables or parameters variations in the values of which will permit the derivation of other explananda which are appropriately different from *E*. (Woodward 1979, 46.)

The *D-N* 'explanation' of the sinking of the eight ball is defective because of its failure as a resource for answering 'what-if-things-had-been-different' questions: it falsely suggests that if Pennsylvania Slims had not put a *blue spot* on the cue ball, she would not have made her winning shot. The more traditional counterexamples would receive similar diagnoses.

If one adheres to a counterfactual or probabilistic theory of causation, the coincidence of our judgments of explanatory relevance with those of counterfactual dependence or statistical relevance strongly suggests that explanatory relevance just is causal relevance. Salmon (1971) seems to have had some hope that an adequate probabilistic theory of causation could be given, so this coincidence of judgments is suggestive of how Salmon himself might have arrived at the conclusion that explanatory relevance is a causal concept; I think that it captures what is plausible in this view today. Salmon later came to reject probabilistic theories of causation.[1] In so doing, he divorced his theory of causation from those considerations that made a causal theory of explanatory relevance plausible in the first place.

5. The Conserved Quantity Theory. Many criticisms of Salmon's (1984) theory have focused on his criteria *MT* and *CI*, claiming that they admit of counterexamples (e.g., Glymour 1982, Kitcher 1989, and Dowe 1992.) Kitcher presents the following example:

> Imagine that a vehicle equipped with skis is sliding on an ice rink and casting a shadow. A projectile is thrown in such a way that it lands at the edge of the shadow with a horizontal component of velocity equal to that of the shadow of the vehicle. Because the projectile lies across the edge there is an immediate distortion of the shadow shape. Moreover, the distortion persists because the projectile retains its position relative to the vehicle (and to its shadow). (Kitcher 1989, 464)

In this example, the moving shadow meets condition *MT*, and so should be countenanced as a causal process, but moving shadows are paradigmatic pseudoprocesses.

The moral that Kitcher draws from such examples is that Salmon's theory of causation is beset with epistemological difficulties. A successful theory of causation ought to account for our ability to acquire causal knowledge. The counterexamples show that *MT* and *CI* are not reliable guides to the discovery of causal structure. More generally, because Salmon's definitions of causal processes and interactions rely so heavily on counterfactuals, and because the truth values of these counterfactuals are not as readily determined by empirical tests as Salmon would have us believe, causal relations become epistemically inaccessible on Salmon's theory (Kitcher 1989, 470–475).

I think that Kitcher is overstating the case here. By drawing out the presuppositions of the first signal principle of special relativity, and by giving numerous illustrations, Salmon has made it abundantly clear that there is a genuine physical distinction between causal processes and interactions and their acausal cousins. The application and tracing of marks often enables us to discover causal processes and interactions, although in these post-positivist times we do not expect there to be any specifiable

---

[1] See (Salmon 1984, chapter 7) for his reasons. Salmon did not explicitly attack Lewis's counterfactual theory in this chapter, although his criticisms encompass it. In particular, he considers 'the method of successive reconditionalization' as an attempt to save probabilistic theories of causality from certain counterexamples (Salmon 1984, pp. 196–202); this is essentially the method that Lewis himself employs to rescue the probabilistic version of his counterfactual theories from similar counterexamples (Lewis 1986, pp. 179–180).

set of observable conditions that is both necessary and sufficient for the presence of causal processes and interactions. The counterexamples presented by Kitcher and others may well undermine *MT* and *CI* as definitions of causal processes and interactions, but it does not follow that they are epistemically inaccessible or otherwise philosophically suspect. While a more accurate characterization of causal processes and interactions would be desirable, the general program of appealing to causal processes and interactions in an attempt to understand the aims of scientific explanation is not seriously undermined by such counterexamples.

Nonetheless, Salmon *is* concerned to provide a more adequate characterization of causal processes and interactions, so he takes these criticisms to heart. He follows Kitcher in pinning the blame for the counterexamples on the counterfactuals involved in the formulation of *MT* and *CI*. Influenced heavily by Dowe (1992), he replaces these definitions with the following counterfactual-free definitions:

1. A causal interaction is an intersection of world-lines which involves exchange of a conserved (invariant) quantity.
2. A causal process is a world-line of an object that transmits a nonzero amount of a conserved (invariant) quantity at each moment of its history (each spacetime point of its trajectory).
3. A process transmits a conserved (invariant) quantity from A to B ($A \neq B$) if it possesses this quantity at A and at B and at every stage of the process between A and B without any interactions in the half-open interval $(A, B]$ that involve an exchange of that particular conserved (invariant) quantity. (Salmon 1994, 303–308, with some minor changes made for clarity.)

A quantity is conserved if, in a closed system, it is constant under translations in time; a quantity is invariant if it is constant under coordinate transformations corresponding to changes in reference frame; and a quantity is exchanged if at least one of the processes involved in an intersection experiences a change in the value of that quantity. (Dowe (1992, 210) offers a slightly more refined definition that accommodates Y- and λ-type interactions as well as X-type interactions.) Definition 1 is taken straight from Dowe (1992) and definition 2 is a modified version of a definition found there. Salmon discusses the merits of formulating the theory in terms of invariant quantities rather than conserved quantities, and tentatively opts for the former.

In light of the arguments of the previous section, it might be expected that the need for counterfactuals is not so easily avoided. Suppose that a shadow is cast on a metal plate that has a uniform nonzero charge density on its surface. The shadow then moves across the plate in such a way that the area of the plate in shadow remains constant. The shadow then possesses a constant quantity of electric charge (a quantity that is both conserved and invariant) as it moves across the plate. The shadow is not participating in any causal interactions as it moves; in particular it is not being bombarded with photons as is the spot of light in a similar example discussed by Salmon (1994, 308). By definition 3, the shadow transmits the charge, and by definition 2, it is a causal process. Note that the older, counterfactual-laden theory correctly rules that this moving shadow is a pseudoprocess, since any (hypothetical) local modification of the charge density would not be transmitted by the shadow. A counterexample to definition 1 results when two such shadows cross and change size—perhaps as a result of some interaction between the objects casting the shadows—thus 'exchanging' electric charge.

It will not do to reject this counterexample on the grounds that the charge possessed by the moving shadow is not *strictly* conserved (because it will not be the case that one electron always enters the shadow precisely when another one leaves it), or on the grounds that it does not really participate in *no* further interactions (even shadows are hit by photons). These considerations would rule out any macroscopic process, such as the speeding bullet discussed by Salmon (1994, 309). Strictly speaking, one causal process dies and a new one is born whenever the bullet interacts with an air molecule and exchanges energy; but for most purposes, we may treat the bullet as a continuous process. It seems that definitions 1, 2 and 3 entitle us to treat the charged shadow similarly.

Another difficulty concerns the choice between formulating the theory in terms of conserved or invariant quantities. It seems to me that conserved quantities are uniquely suited to the goal of purging counterfactuals. Consider Salmon's old definition of causal interaction, *CI*. According to this definition, an intersection of processes is a causal interaction only if each process undergoes a change that *would not have taken place* without the intersection. The counterfactual clause is needed: the *positions* of two light pulses that cross undergo change, but this is a paradigm example of an intersection that is *not* a causal interaction. *CI* avoids counterexample because the positions would have changed in any event. It is not clear how a restriction to invariant quantities avoids this need for recourse to counterfactuals. Any quantity that is not conserved may undergo uncaused change, making possible spurious 'exchanges' like that described above. The advantage of conserved quantities is that *any* change in a conserved quantity is one that would *not* have happened otherwise.

This last observation also points to a difficulty for the conserved quantity formulation of the theory, however. Suppose it be asked how we are to define 'conserved quantity'. A conserved quantity is one that remains constant through time in a closed system, but what is a closed system but a system that does not engage in any causal interactions? In other words, is not the concept of conserved quantity to be explicated in terms of the concept of causal interaction, rather than the other way around? It seems to me that conservation laws *presuppose* the notion of causal interaction in much the same way that the first signal principle of special relativity presupposes the notion of a causal process. Without the restriction to causal processes, the first signal principle is false; without the exclusion of causal interactions, the law of conservation of electric charge is false.

I suggest that the conserved quantity theory is best viewed as augmenting rather than replacing the mark-transmission theory. Neither theory provides a reductive analysis of the concepts of causal process and interaction, and neither provides infallible rules for detecting causal processes and interactions. Rather, each provides *guidelines* for recognizing causal processes and interactions, as well as reasons for thinking that these concepts are presupposed by physical science.

Let us put these considerations aside, however, and take the new theory at face value. The question of central concern is whether it avoids the difficulties raised in the previous section. The answer is "no": the arguments of the previous section may be repeated *mutatis mutandis*. In describing the location of some explanandum within a network of causal processes and interactions, one implies only that *some* conserved quantity is transmitted by each process, and that *some* conserved quantity is exchanged during each interaction. Even if we are entitled, when giving explanations, to appeal to the quantities that are in fact transmitted and exchanged, the explanatorily relevant properties will not be singled out. Consider again the example of Pennsylvania Slims' winning shot. The moving cue ball transmits many conserved quantities: linear momentum, angular momentum, and electric charge, for example. (Actually, the moving cue ball will be continually exchanging these quantities with the pool table, but we may treat the continuous transmission of these quantities as a good approximation.) Of these, only linear momentum is explanatorily relevant to the final location of the eight ball. (It is left as an exercise for the reader to rewrite the last three sentences in terms of *invariant* quantities.) And this is the example that is superficially most congenial to Salmon's approach: the prospects for explaining John Jones's failure to become pregnant in terms of the exchange and transmission of conserved quantities never looked bright.

The concepts of causal process and interaction are physically legitimate and philosophically interesting regardless of whether either of Salmon's theories provides a definitive characterization of them. But these concepts do not provide the resources to explicate the notion of explanatory relevance that Hempel's models failed to capture.

6. The S-R Basis. The target of my attack has been a theory of explanation consisting of the conjunction of the following theses: (1) causal explanations provide descriptions of the causal nexus; and (2) the causal nexus is nothing more than a network of physical processes and interactions described (albeit imperfectly) by (2a) *MT* and *CI*, or (2b) definitions 1 through 3. But are causal processes and interactions really the only explanatory resources available on Salmon's view? Salmon (1984) incorporates part of his earlier (1971) *Statistical-Relevance* (*S-R*) theory of explanation; in the later book, Salmon still maintains that explanations have an "*S-R* basis":

> . . . explanation is a two-tiered affair. At the most basic level, it is necessary, for purposes of explanation, to subsume the event-to-be-explained under an appropriate set of statistical relevance relations, much as was required under the *S-R* model. (1984, 22)

The *S-R* model did offer an account of explanatory relevance; the inclusion of the *S-R* basis in Salmon's (1984) theory of explanation suggests that he may simply co-opt this treatment.

Salmon does not take statistical relevance to be constitutive of causal relevance: Salmon (1984) repudiates probabilistic theories of causation, and Salmon (1994) reiterates this. So if Salmon is to retain the treatment of explanatory relevance in terms of statistical relevance, he must either: (i) admit that explanatory relevance is not a causal concept; or (ii) establish a role for the *S-R* basis within genuinely causal explanations. Salmon would clearly prefer to follow the second route, else he would not have felt the need to supplement the *S-R* model with a theory of causation. Unfortunately, the explanatory import of the *S-R* basis within Salmon's (1984) theory is not clear. In some passages, Salmon seems to suggest that the *S-R* basis is like an explanatory stepping stone, to be itself explained in terms of causal structure. For example, the passage just cited continues:

> At the second level, it seems to me, the statistical relevance relations that are invoked at the first level must be explained in terms of *causal* relations. (1984, 22; see also the similar passage on p. 261.)

Other passages suggest that probability relations only supply evidence for the causal relations that are genuinely explanatory:

> The explanatory relevance of statistical relevance relations is indirect. Their fundamental import lies in the fact . . . that they constitute evidence for causal relations. (1984, 192; see also note 10, p. 34, and p. 265, where Salmon alludes approvingly to others who have expressed this view.)

Pending further clarification, it is difficult to assess the prospects for providing a causal account of explanatory relevance by means of the *S-R* basis.

Here is a very programmatic suggestion. In Hitchcock (1995), I show that objections to probabilistic theories of causation of the sort leveled by Salmon may be avoided. But even those who defend probabilistic theories of causation often claim that it is not possible to provide a reductive analysis of causation in terms of probability alone. Typically, the goal of the enterprise has been seen as one of imposing probabilistic *constraints* upon an antecedently given causal relation. (See, for example, Eells 1991.) I have argued (1993) that we instead conceive of a probabilistic theory of causation as providing a *taxonomy* of the primitive relation of causal relevance. Either way, it is quite possible that the causal processes and interactions described by Salmon provide the foundation for this primitive relation. If this is correct, it may be possible to collapse Salmon's two explanatory tiers so that causal and explanatory relevance relations are constituted jointly by relations of probabilistic relevance *and* physical connections. The proposed marriage of probabilistic and process theories of causation is one that merits further exploration.

7. The New Mechanical Philosophy. Salmon, following up on a remark made in Glymour (1982), refers to his account of scientific explanation as 'the mechanical philosophy' (Salmon 1984, 278–279). Glymour was alluding to Salmon's apparent commitment to action by contact, but I think that the similarities between Salmon's mechanism and that of his seventeenth-century predecessors are more profound than has previously been realized. Central to the world view of the seventeenth-century mechanists was the distinction between primary and secondary qualities. The primary qualities were size, shape, position, motion and duration (this list is Descartes'; he sometimes includes number). These were contrasted with the secondary qualities of heat, taste, odor, sound, and color. The primary qualities were those properties of objects that could be characterized geometrically. According to the mechanists, all physical phenomena (in contrast with mental phenomena) were ultimately to be explained in terms of the primary qualities of objects. Historian of science Eduard Dijksterhuis writes:

> The only properties recognized as explanatory principles were the size, the shape, and the state of motion of corpuscles, supplemented by characteristics of their aggregates that could also be defined geometrically (1961, 432.)

The appeal of the mechanical philosophy lay, in part, in the austerity of its explanatory store: by claiming that all phenomena could be explained in terms of the geometric properties of physical objects, the mechanical philosophy offered the promise of a simple, unified theory of nature. But its source of appeal was also its undoing: the explanatory store was simply too impoverished to account for the phenomena of the physical world.

According to the new mechanism, to explain a phenomenon is to show how it fits into a causal nexus. Salmon's theory is appealing for much the same reason that the world view of his seventeenth-century forebears was: it is ontologically austere. The causal nexus consists only of physical processes and interactions of a sort that we are already committed to, both in everyday life and in physical science. The theory is refreshingly free of metaphysics: there is no mysterious cement of the sort for which Hume sought in vain—and in the later theory, modality is banished.

But alas, the new mechanical philosophy must go the way of the old. In the new mechanism, relations between the properties or quantities possessed by a process play a role in determining whether that process is causal or not, but these relations are not what we point to when we locate an explanandum within its causal nexus. What, then, is left to do the explaining? Only the naked processes and interactions, only a network of spacetime worms such as that depicted in figure 1. Such an explanation characterizes only a particular (albeit intricately fibrillated) region of spacetime as relevant to the outcome. In the new mechanical philosophy, as in the old, the explanatory store contains nought but geometric properties.

REFERENCES

Dijksterhuis, E. J. (1961), *The Mechanization of the World Picture*. Princeton: Princeton University Press.

Dowe, P. (1992), "Wesley Salmon's Process Causality and the Conserved Quantity Theory", *Philosophy of Science 59*: 195–216.

Dray, W. (1957), *Laws and Explanation in History*. Oxford: Oxford University Press.

Eells, E. (1991), *Probabilistic Causality*. Cambridge, England: Cambridge University Press.

Glymour, C. (1982), "Causal Inference and Causal Explanation", in R. McLaughlin (ed.), *What? Where? When? Why?* Dordrecht: Reidel, pp. 179–191.

Hempel, C. G. (1965), "Aspects of Scientific Explanation", in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, pp. 331–496.

——— and Oppenheim, P. (1948), "Studies in the Logic of Explanation", *Philosophy of Science 15*: 135–175. (Reprinted with added postscript in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*.)

Hitchcock, C. R. (1993), "A Generalized Probabilistic Theory of Causal Relevance", *Synthese 97*: 335–364.

———. (1995), "The Mishap at Reichenbach Fall: Singular vs. General Causation", *Philosophical Studies 78*, June (in press).

Kitcher, P. (1989), "Explanatory Unification and the Causal Structure of the World", in Kitcher and Salmon (1989), pp. 410–505.

Kitcher, P. and Salmon, W. C. (eds.) (1989), *Scientific Explanation*. Minneapolis: University of Minnesota Press.

Kyburg, H. (1965), "Discussion: Salmon's Paper", *Philosophy of Science 32*: 147–151.

Lewis, D. K. (1986), "Causation" and "Postscripts to 'Causation' ", in *Philosophical Papers, Volume II*. Oxford: Oxford University Press, pp. 159–213.

Salmon, W. C. (1965), "The Status of Prior Probabilities in Statistical Explanation", *Philosophy of Science 32*: 137–146.

———. (1971), "Statistical Explanation", in Salmon et al., *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press, pp. 29–87.

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

———. (1994), "Causality without Counterfactuals", *Philosophy of Science 61*: 297–312.

Skyrms, B. (1984), "EPR: Lessons for Metaphysics", in P. French, T. Uehling, and H. Wettstein (eds.), *Midwest Studies in Philosophy IX: Causation and Causal Theories*. Minneapolis, University of Minnesota Press, pp. 245–255.

Woodward, J. (1979), "Scientific Explanation", *British Journal for the Philosophy of Science 30*: 41–67.

———. (1984), "A Theory of Singular Causal Explanation", *Erkenntnis 21*: 231–62.

———. (1989), "The Causal Mechanical Model of Explanation", in Kitcher and Salmon (1989), pp. 357–383.