7

Explanatory Realism, Causal Realism, and Explanatory Exclusion

I

Explaining is an epistemological activity, and "having" an explanation is, like knowing, an epistemological accomplishment. To be in need of an explanation is to be in an epistemologically imperfect state, and we look for an explanation in an attempt to remove that imperfection and thereby improve our epistemic situation. If we think in terms of the traditional divide between knowledge and reality known, explanations lie on the side of knowledge—on the side of the "subjective" rather than that of the "objective," on the side of "representation" rather than that of reality represented. Our explanations are part of our knowledge of the world.

Knowledge implies truth: we cannot know something that is not the case. On a realist view of knowledge, every bit of knowledge has an objective counterpart, the thing that is known which is itself not part of knowledge—at least, not part of that particular bit of knowledge. But exactly what is it that we know when we have an explanation? Exactly in what does *explanatory knowledge* consist? If explanations constitute knowledge, it makes sense to ask, for each explanation that we "have," exactly what it is that we know in virtue of having that explanation. And when we gain a new explanation, precisely what change takes place in our body of knowledge? We usually think of knowledge as consisting of a set of "propositions," thought to represent

I am indebted to David Benfield, Paul Boghossian, Brian McLaughlin, Joseph Mendola, and Michael Resnik for discussions of some of the issues taken up in this paper.

"facts" of the world. These propositions are discrete items, although they form a complex network of logical and evidential connections. How are we to represent explanatory knowledge within such a picture? Where do we locate explanations in a scheme of propositions?

We can think of an explanation as a complex of propositions or statements divisible into two parts, *explanans* and *explanandum proposition*.[1] Since explanations can take a variety of linguistic forms, this division is rough; in particular, it is not to be taken to imply that an explanation is an argument or inference, with the explanans as premise and the explanandum as conclusion. Let us focus on explanations of individual events. Such explanations typically explain an event (why a given event occurred) by reference to another event (or set of events). Let *E* be the explanandum, to the effect that a certain event *e* occurred. Let *C* be an explanans for this explanandum. *C*, let us assume, is the statement that event *c* occurred. Suppose then that we "have" this explanation—that is, *C* and *E* are related as explanans to explanandum in our body of knowledge (call this the "explanans relation"). What is the relationship between events *c* and *e*?

What I want to call *explanatory realism* takes the following position: *C* is an explanans for *E* *in virtue of* the fact that *c* bears to *e* some determinate objective relation *R*. Let us call *R*, whatever it is, an "explanatory relation." (The *explanans* relation relates propositions or statements; the *explanatory* relation relates events or facts in the world.) The explanatory relation is an objective relation among events that, as we might say, "ground" the explanans relation, and constitutes its "objective correlate." On the realist view, our explanations are "correct" or "true" if they depict these relations correctly, just as our propositions or beliefs are true if they correctly depict objective facts; and explanations could be more or less "accurate" according to how accurately they depict these relations. Thus, that *c* is related by explanatory relation *R* to *e* is the "content" of the explanation consisting of *C* and *E*; it is what the explanation "says."

Although the attribution of truth or correctness to explanations is essential to explanatory realism, it by itself is not sufficient; those who reject explanatory realism in our sense, too, can speak of the truth or falsity of an explanation—for example, in the sense that propositions constituting the

[1] I shall sometimes use "explanandum" as short for "explanandum proposition (or statement)." This should cause no confusion.

explanans are all true. What matters to realism is that the truth of an explanation requires an *objective relationship* between the events involved. By an "objective relation," I have in mind a relation that at least meets the following condition: that it is instantiated does not entail anything about the existence or nonexistence of any intentional psychological state—in particular, an epistemological or doxastic state—except, of course, when it is instantiated by such states. I am not suggesting that the explanatory relation holding for events is all there is to explanations, or to the explanans relation. Just as knowledge requires more than truth, explanations presumably must meet further requirements ("internal" conditions—perhaps logical and epistemic ones), although exactly what these are does not concern us here.

What could such an *R* be in virtue of which an event is correctly cited in the explanation of another? The obvious first thought is this: *R* is the *causal relation*. Perhaps there are noncausal explanations of individual events; however, few will deny that the causal relation is at least one important special case of *R*. And there are those who hold that the causal relation is the only explanatory relation—at least the principal one.[2]

Explanatory irrealism, on the other hand, would be the view that the relation of being an explanans for, as it relates *C* and *E* within our epistemic corpus, is not, and need not be, "grounded" in any objective relation between events *c* and *e*. It is solely a matter of some "internal" relationship between items of knowledge. Perhaps, there are logical, conceptual, or epistemic relationships among propositions in virtue of which one proposition constitutes an explanans for another, and when that happens, we could speak of the events represented as being related by an explanatory relation. That is, given the explanans relation over propositions, a relation over the events they represent could be defined: *c* explains (is related by *R* to) *e* just in case *C* is an explanans for *E*. But an *R* so defined would fail to be an objective relation, as required by realism, for it would depend crucially on what goes on within our body of knowledge and belief.

[2] e.g., Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton, NJ, 1984); David Lewis, "Causal Explanation," in *Philosophical Papers*, vol. 2 (Oxford, 1986). However, there are relations other than causation one might want to consider: e.g., the relation of supervenience, the micro-reductive relation. Whether or not these possible explanatory relations require the same explanandum as the causal relation is another question; see Robert Cummins's distinction between "explanation by subsumption" and "explanation by analysis" in his *The Nature of Psychological Explanation* (Cambridge, Mass., 1983), chap. 1. See also Peter Achinstein, "A Type of Non-Causal Explanation," *Midwest Studies in Philosophy* 9 (1984): 221–243.

In the following passage Wesley Salmon gives a clear and forceful expression to the realist view of explanation:

> We need not object to [the purely psychological conception of explanation] merely on the ground that people often invoke false beliefs and feel comfortable with the 'explanation' thus provided. . . . We can, quite consistently with this approach, insist that adequate explanations must rest upon *true* explanatory bases. Nor need we object on the ground that supernatural 'explanations' are often psychologically appealing. Again, we can insist that the explanation be grounded in *scientific* fact. Even with those restrictions, however, the view that scientific explanation consists in release from psychological uneasiness is unacceptable for two reasons. First, we must surely require that there be some sort of *objective* relationship between the explanatory facts and the fact-to-be-explained.[3]

However, merely to hold that C is an explanans for E just in case c is a cause of e is not necessarily to espouse explanatory realism. Whether that is so depends on one's conception of causation. Consider, for example, Hanson, who writes:

> The primary reason for referring to the cause of x is to explain x. There are as many causes of x as there are explanations of x.[4]
>
> Causes certainly are connected with effects; but this is because our theories connect them, not because the world is held together by cosmic glue. The world *may* be glued together by imponderables, but that is irrelevant for understanding causal explanation. The notions 'the cause x' and 'the effect y' are intelligible only against a pattern of theory, namely one which puts guarantees on inference from x to y.[5]

For Hanson, causal relations essentially depend on an appropriate conceptual interlocking of our descriptions as provided by the theories we accept. He makes it evident, in the quoted passages, that he views the causal relation between x and y as derivative from an inferential relation from x to y, and the inferential relation as intimately associated with explanation; it is also

[3] *Scientific Explanation and the Causal Structure of the World*, p. 13 (emphasis in the original). Explanatory realism, as I have characterized it, appears closely related to what Salmon calls "the ontic model" of scientific explanation.

[4] Norwood Russell Hanson, *Patterns of Discovery* (Cambridge, 1958), p. 54.

[5] Ibid. 64. See, for a view similar to Hanson's but worked out in greater detail, William Ruddick, "Causal Connection," *Synthese* 18 (1968): 46–67.

evident that he does not take the dependence of causation on inference and explanation to be merely epistemological. If one accepts this view of causation and causal explanation, there is nothing realist about the position that causal explanations hold just in case the causal relation holds. For causal relations, on such an approach, depend on inferential-explanatory connections which are primary and more basic.

More generally, if one wants to *analyze* causation itself in terms of explanation,[6] one would be rejecting explanatory realism—unless one could identify an objective relation other than causation as the explanatory relation. But what could such a relation be? One might wish to propose the *nomological* relation as a candidate. The idea is this: that two events, *c* and *e*, are "subsumed under," or "instantiate," an appropriate law is the objective correlate of the explanans relation for *C* and *E*. Giving an account of "subsumption under a law" without presupposing causal notions is not an easy task, but let us not press this issue.[7] The point to consider is how we understand the notion of "law." If a law is taken as "mere Humean constant conjunction," with no modal or subjunctive force intimating some tie of "necessitation," this approach would give us realism. But it is highly dubious that a conception of an explanatory relation based on such a notion of "law" could provide a basis for an adequate account of explanation; it is even more dubious that an analysis of causation based on such a conception of explanation will come close to capturing our concept of causation. On the other hand, if laws are endowed with sufficiently strong modal force, it is doubtful whether the nomological relation will be distinguishable, in any meaningful way, from the causal relation.[8] Indeed, the nomological account of causation is one of the more influential approaches to the analysis of the causal relation. An analysis of causation in terms of a conception of explanation that in turn is based on the nomological relation as the explanatory relation will essentially be just a nomological analysis, possibly with some psycho-epistemological embellishments. It would be difficult to see why

[6] See, for example, Michael Scriven, "Causation as Explanation," *Noûs* 9 (1975): 3–16. I discuss Scriven's account in "Causes as Explanations: A Critique," *Theory and Decision* 13 (1981): 293–309. Some of the present material has been drawn from this paper.

[7] For some general difficulties in explaining "subsumption under a law," see Donald Davidson, "Causal Relations," *Journal of Philosophy* 64 (1967): 691–703, and my "Causation, Nomic Subsumption, and the Concept of Event," *Journal of Philosophy* 70 (1973): 217–236.

[8] Where the nomological and the causal relation do not match up, the former also fails to yield the explanatory relation.

one should not just go for a direct nomological analysis of causation, and use the causal relation as one's explanatory relation.

We must conclude that any attempt to analyze causation as explanation will result in a form of explanatory irrealism. For an analysis of causation to be a genuine explanatory analysis, the concept of explanation assumed as the basis of analysis must be a robustly epistemological and psychological notion whose core is constituted by such notions as understanding and intelligibility, not some pale, formal reconstruction of it. If, for example, the Hempelian deductive-nomological conception of explanation is used to explain causation, the result is not a genuine explanatory analysis of causation but rather the old standby, the nomological-subsumptive, or quasi-Humean, analysis. Thus, on a real explanatory approach to causation, causation will turn out to be a nonobjective psycho-epistemological relation and, therefore, fail to serve as an objective correlate of the explanans relation.

II

It is plausible to conclude, therefore, that explanatory realism requires the causal relation as an explanatory relation. As I said, we may leave open the question whether the causal relation is the only explanatory relation. But at least this much is certain: in both everyday and most scientific contexts,[9] explanations of individual events are predominantly *causal explanations* in the sense that the events cited in the explanation of an event are its causes and, further, their explanatory efficacy is thought to stem from their causal status. And when each of a class of events can be given a similar causal explanation, we may have a causal explanation of a regularity. We shall in this paper focus exclusively on the causal relation as our explanatory relation; our general metaphysical points should be valid, *mutatis mutandis*, for other explanatory relations if any exist. Explanatory realism says this about causal explanations: a causal explanation of event *e* in terms of event *c* ("*e* occurred because *c* caused it") is *correct*, or *true*, just in case *c* did as a matter of objective fact cause *e*. That the causal relation holds between the two events constitutes the "factual content" of the explanation. This may sound obvious and trivial.

[9] By the qualification "most," I intend to leave out consideration of what some tell us goes on at the deepest and most abstract levels of theoretical physics.

Perhaps it sounds obvious only because we take explanatory realism for granted. But it certainly is not trivial. It requires, for its intended realist purposes, that causality itself be an objective feature of reality. This doctrine, which we may call *causal realism*, has not gone unchallenged. Hume's celebrated critique of "necessary connection" as an objective relation characterizing events themselves was perhaps the first—clearly the most influential—expression of a systematically articulated, irrealist position on causation. He wrote, "Upon the whole, necessity is something that exists in the mind, not in objects."[10] Hume well understood the causal realist's sentiments:

> But though this be the only reasonable account we can give of necessity, the contrary notion is so riveted in the mind from the principles above-mentioned, that I doubt not but my sentiments will be treated by many as extravagant and ridiculous. What! the efficacy of causes lie in the determination of mind! As if causes did not operate entirely independent of the mind, and would not continue their operation, even though there was no mind existent to contemplate them, or reason concerning them. Thought may well depend on causes for its operation, but not causes on thought. This is to reverse the order of nature, and make that secondary, which is really primary.[11]

Hume was understanding, but in the end dismissive:

> I can only reply to all these arguments that the case here is much the same as if a blind man should pretend to find a great many absurdities in the supposition that the colour of scarlet is not the same with the sound of a trumpet, nor light the same with solidity. If we have really no idea of a power or efficacy in any object, or of any real connection betwixt causes and effects, it will be to little purpose that an efficacy is necessary in all operations.[12]

Hume regarded the other ingredients he identified in the causal relation, namely temporal precedence, spatiotemporal contiguity (or connectability), and constant conjunction, as objective and mind-independent features of causally connected events;[13] but evidently he thought that necessitation, too,

[10] *A Treatise of Human Nature*, ed. L. A. Selby-Bigge (Oxford, 1888), p. 165.

[11] Ibid. 167.

[12] Ibid. 168.

[13] For discussions of this and other matters concerning Hume on causation and necessity, see Barry Stroud, *Hume* (London, 1977), chaps. 3, 4, and Tom L. Beauchamp and Alexander Rosenberg, *Hume and the Problem of Causation* (New York and Oxford, 1981), especially chap. 1.

was an essential element in our philosophically unenlightened (by his light) concept of causation. Most philosophers will now agree that an idea of causation devoid of some notion of necessitation is not *our* idea of causation—perhaps not an idea of causation at all. According to most conceptions of causation now current, at any rate, Hume was a causal irrealist par excellence.

Hume, our original causal irrealist, had some illustrious followers. Russell ridiculed causation as "a relic of a bygone age," recommending the "extrusion" of the word "cause" from the philosophical vocabulary;[14] Wittgenstein said, "Belief in causal nexus is superstition."[15] The positivist-inspired suspicion of modalities, counterfactuals, and the like, which characterized much of analytic philosophy during the first two-thirds of this century, is of a piece with Hume's causal irrealism in their fundamental philosophical motivation, and it seems that many prominent philosophers in the analytic tradition during this period consciously avoided serious discussion of causality, making little use of it in their philosophical work.[16] More recently, Hilary Putnam has attacked the idea of "non-Humean causation" as a physically real relation.[17] I think it is more difficult than one might at first suppose to find philosophers who have consciously advocated an unambiguously realist conception of causality.[18]

According to causal realism, therefore, causal connections hold independently of anyone's intentional states—in particular, epistemological or doxastic states—except, of course, when the causal connections concern such states. The realist believes, as Hume observes in the quotation above, that causal relations—the same ones—would hold even if there were no conscious beings to "contemplate them, or reason concerning them." This means that according to causal realism every event has a *unique and determinate causal*

[14] Bertrand Russell, "On the Notion of Cause," *Proceedings of the Aristotelian Society* 13 (1913): 1–26.

[15] Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (London, 1922), 5.1361. However, Wittgenstein may have had in mind by "causal nexus" something much stronger than what we would now understand by "causal necessity."

[16] C. J. Ducasse and Hans Reichenbach were among the exceptions.

[17] "Is the Causal Structure of the Physical Itself Something Physical?," *Midwest Studies in Philosophy* 9 (1984): 3–16.

[18] Some possibilities among recent writers: Salmon, *Scientific Explanation and the Causal Structure of the World*; J. L. Mackie, *The Cement of the Universe* (Oxford, 1974), especially chap. 8. Quine seems to have studiously avoided discussing causation or making use of it in his philosophical work. The uses to which Donald Davidson has put the concept of causation indicate a realist attitude; consider, for example, his commitment to an event ontology and his causal criterion of event individuation in his *Essays on Actions and Events* (Oxford, 1980). But he may reject the terms in which I have formulated the positions.

history whose character is entirely independent of our representation of it. We may come to know bits and pieces of an event's causal history, but whether we do, or to what extent we do, and what conceptual apparatus is used to depict it, do not in any way affect the causal relations in which events stand to other events. This entails that the existence and character of events themselves must be an objective and determinate fact; that is, causal realism makes sense only in the context of global realism.

Earlier, we raised the question of how explanatory knowledge is represented in our body of knowledge—that is, what it is that we know when we have an explanation of an event. The explanatory realist appears to have a simple answer: To "have an explanation" of event *e* in terms of event *c* is to know, or somehow represent, that *c* caused *e*; that is, explanatory knowledge is causal knowledge, and explanations of individual events are represented by singular causal propositions. Thus, explanatory knowledge is propositional knowledge of a certain kind, and to gain an explanation of an event is to learn *a further fact about that event*.

But is there an alternative to representing explanations as additional bits of propositional knowledge? Isn't explanatory knowledge a kind of knowledge, and isn't all knowledge, in an epistemologically relevant sense, a matter of knowing *that*? Although I do not know whether anyone has held a view like this, it is possible to hold, I think, that explanations are essentially a matter of how a body of knowledge is organized or systematized—a matter of there being certain appropriate patterns of coherence among items of knowledge. That is, to "have" an explanation of why *E* in terms of *C*—that is, to "have" *C* as an explanans for *E*—is simply for the two propositions *C* and *E* to be appropriately related within our epistemic corpus; it is not a matter of there being a further proposition within it. According to this view, therefore, explanatory knowledge supervenes on nonexplanatory knowledge: if you and I know exactly the same first-order, factual propositions (roughly, propositions that can serve as explananda and elements of an explanans), we would share the same explanations. Various considerations might lead us to qualify this conclusion; for example, one might construe the notion of "having an explanation" in such a way as to require the subject's awareness that the explanans is appropriately related to the explanandum. Thus, one might want to suggest that the presence of the two propositions in our body of knowledge is not enough, even if they in fact instantiate a required explanatory pattern, and insist that

we must somehow mentally "bring them together" and "see" that they do so. Caution is required, however; pursuing this line may take one back to the propositional view of explanatory knowledge. At any rate, the nonpropositional, "pattern" view of explanatory knowledge differs from the propositional view on the following point: gaining a new explanation, on the pattern view but not on the propositional view, does not necessarily involve acquiring new information about facts of the world.

It seems clear that explanatory realism leads to the propositional view of explanatory knowledge; it makes "having" an explanation a matter of knowing a certain proposition to be true. On the other hand, explanatory irrealism, although it has an affinity for the pattern view, is not committed to it; it appears consistent with the propositional view. One might hold, for example, that a certain conceptual-epistemic relation between an explanandum and its explanans is what is fundamentally constitutive of an explaining relation, there being no independent objective relation characterizing the events represented by the explanans and the explanandum that grounds it,[19] but that "having" an explanation *is* a matter of *knowing that* this relationship does in fact hold for the explanans-explanandum pair. This, however, may not be a plausible view; it is naturally construed as requiring anyone who "has" any explanation of anything at all to know what the explanans relation is, something that few philosophers would confidently claim to know. In any case, those who find the propositional view of explanatory knowledge too simplistic, or otherwise unpalatable, would have to settle for explanatory irrealism; explanatory realism is not an option for them.

What difference does the choice between explanatory realism and irrealism make? We have already seen that explanatory realism plausibly entails causal realism. Does explanatory irrealism entail causal irrealism? There evidently is no strict inconsistency in holding both explanatory irrealism and causal realism. However, the combination seems somewhat incongruous and difficult to motivate: though acknowledging causation as a genuine relation in the world, the position denies it any essential role in explanation, severing the intuitive and natural tie between causality and explanation.[20] What, then, would be the point of the causal relation? The concept of

[19] Recall our earlier discussion of Hanson; he seems to have held a view like this.

[20] Peter Achinstein's views in *The Nature of Explanation* (New York and Oxford, 1983), chap. 7, seem to approximate this position.

causation, of course, has many roles to play, but it seems that its explanatory role is a central one, being closely tied to its other important roles. I think that the combination of explanatory irrealism and causal realism, though logically consistent, is not a plausible position.

We have also seen that explanatory realism entails the propositional account of explanatory knowledge, whereas explanatory irrealism, again, seems consistent with each of the two alternatives, the propositional view and the nonpropositional, pattern view. I think that the issue of causal realism versus irrealism and that concerning the nature of explanatory knowledge are significant issues, both interesting in themselves and important in what they imply for other philosophical problems. Problems about what explanatory knowledge consists in—that is, what "understanding" something amounts to—have been almost entirely neglected within traditional epistemology; this is surprising in view of the centrality of explanation in philosophy of science, which, by and large, is the epistemology of scientific knowledge. The issue of causal realism is obviously important: whether causal relations are real and objective, or mere projections of the cognizing mind, is an issue that directly affects the significance of causation within both science and philosophy. If it is an objective relation characterizing physical events in the world, is it physically reducible, or physically based in some sense, as we expect of other physical properties and relations? If not, what accounts for its special status? Which of the special sciences are responsible for investigating the properties of the causal relation itself?[21]

As for the philosophical implications of the choice between causal realism and irrealism, it is an interesting question, for example, whether any of the so-called causal theories (of perception, memory, knowledge, action, event-identity, reference, time, persistence, properties, and no doubt many others) will be able to retain, under an irrealist conception of causation, what plausibility it enjoys. It is also an interesting question whether a substantive version of global realism can be combined with causal irrealism. I suspect that if all causal facts are taken away from the world, not much of interest may remain—the world would become so impoverished, a pale imitation of a world, that we may not care much whether it is real or only "ideal." (If all those "causal theories" mentioned above are true, a world devoid of causal relations would be one in which there is no perception,

[21] See Putnam, "Is the Causal Structure of the Physical Itself Something Physical."

no knowledge, no naming or referring, no intentional action, no time, no persisting object, and none of the rest. It would also be a world in which there are no killings, no breakings, no pushings or pullings, and so on.)

Some may consider it a disadvantage of explanatory realism that it comes only in a package with causal realism, whereas explanatory irrealism can in principle be purchased separately. However, others may consider that an advantage: causal realism gives more content to explanatory realism, and as a result explanatory realism can do work that its rival cannot. Moreover, there is a certain satisfying unity in the combination of explanatory and causal realism. In any case, there seems to be some incongruity, as we saw, in combining explanatory irrealism and causal realism, so that an explanatory irrealist may in effect have no real choice but to embrace causal irrealism as well. In what follows, I will explore the implications of the realist view of explanation for the issue of "explanatory exclusion" and the irrealist (or "internalist") implications of the Hempelian inferential view of explanation.

III

I have argued elsewhere[22] that proffered explanations of a single event, with mutually consistent explanantia, can exclude one another in the following sense: there can be no more than a single *complete* and *independent* explanation of any one event, and we may not accept two (or more) explanations of a single event unless we know, or have reason to believe, that they are appropriately related—that is, related in such a way that one of the explanations is either not complete in itself or dependent on the other. This constraint on explanations, which we may call *the principle of explanatory exclusion*, has two clauses: the first is about the *existence* of explanations, the second about the *acceptance* of explanations. The first clause, I shall argue, can be seen as a plausible thesis if we assume explanatory realism. We shall not discuss the second clause here.

Suppose, then, that each of C_1 and C_2 is claimed to be a causal explanans for E. Let c_1, c_2, and e be the events represented by C_1, C_2, and E. According to explanatory realism, it follows that c_1 caused e and also that c_2 caused e. How are we to understand this situation? There are various possibilities:

[22] "Mechanism, Purpose, and Explanatory Exclusion," *Philosophical Perspectives* 3 (1989): 77–108.

(1) It turns out that $c_1 = c_2$. A single event is picked out by nonequivalent descriptions. Here, there is in reality only one pair of events related by the explanatory relation (that is, the causal relation), and this gives sense to the claim that there is, here, one explanation, not two. The exclusion principle makes sense only if a criterion of individuation is assumed for explanations—that is, only if we can make sense of "same" and "different" as applied to explanations. Now, explanatory realism yields a natural way of individuating explanations: explanations are individuated in terms of the events related by the explanatory relation (the causal relation, for explanations of events).[23] For on realism it is the objective relationship between events that ultimately grounds explanations and constitutes their objective content. This provides us with a basis for regarding explanations that appeal to the same events standing in the same relation as giving, or stating, one explanation, not two—just as two inequivalent descriptions can represent the same fact. Thus, on explanatory realism, we can make good sense of the idea that logically inequivalent explanations can represent the same explanatory relation, and therefore state the same explanation. To the explanatory irrealist, this way of individuating explanations would be unmotivated: explanations would be more appropriately individuated in terms of descriptions or propositions and their internal logical, conceptual, and epistemic relationships. Nothing needs to prevent the explanatory realist from accepting this "internal" individuation criterion as well, as defining *another* useful sense in which we can count explanations. The point is only that explanatory realism motivates an "objective" individuation of explanations, which is both intuitively plausible and well suited for the exclusion principle.

(2) c_1 is reducible to, or supervenient on, c_2. This sort of relationship might obtain, for example, on some accounts of the mind-body relation, which, though eschewing an outright psychophysical identification, nonetheless recognizes the reductive or supervenient dependency of the mental on the physical. In such a case, the causal relation involving the supervenient or reduced event must itself be thought of as supervenient or reducible to

[23] If relations other than the causal relation can serve as explanatory relation, they can also be considered as a basis for individuation; however, that probably would be redundant. It is unlikely that when the explanatory relation is different, exactly the same events would be involved.

the causal or nomological relation involving the "base" event.[24] In this sense, the two explanations are not independent; for the one involving the reduced causal (that is, explanatory) relation is dependent on the one representing the "base" causal relation. This, again, is an example of realist thinking: dependency between explanations is understood in terms of the dependency between the objective explanatory relations that they represent.

(3) c_1 and c_2 are only partial causes, being constituents in a single sufficient set of causal conditions. Example: You push the stalled car and I pull it, and the car moves. In this case, neither explanation is complete: each gives only a partial picture of the causal conditions that made up a sufficient cause of the effect. This sense of *explanatory completeness*, understood in terms of *sufficient cause*, is again entirely natural within the realist picture. For, according to the realist view, the causal relation between events constitutes the objective correlate, or content, of the explanans relation; where a particular causal relation gives us a cause event that is only a partial cause, or one among the many constituents of a sufficient cause, the corresponding explanans, too, can be thought to be only partial and incomplete. Conversely, when the causal relation provides a sufficient cause, the explanans can also be said to be complete and sufficient. The realist scheme also yields a more global sense of "complete explanation," one in which a complete explanation of an event specifies its entire causal history in every detail (as we noted earlier, under explanatory realism each event has a unique determinate causal history). This is an idealized sense of completeness, and no explanation can be complete in that sense (the notion of an ideally complete explanation, however, may be useful in explicating the concept of explanation).[25] Obviously, in this idealized sense there is at most one complete explanation of any given event; again, though obvious and uninteresting, this is not trivial, unless causal realism is trivial.

(4) c_1 and c_2 are different links in the same causal chain leading to event e. But they are not independent: the later event is, then, causally dependent on

[24] For further discussion see my "Epiphenomenal and Supervenient Causation," *Midwest Studies in Philosophy* 9 (1984): 257–270. I believe that the case in which c_1 "generates" c_2 in Alvin Goldman's sense (see his *A Theory of Human Action* (Englewood Cliffs, NJ, 1970)) can be handled in a similar way, although the details may have to be somewhat different.

[25] Compare Peter Railton's notions of "ideal explanatory text" and "ideal causal D-N text" in his "Probability, Explanation, and Information," *Synthese* 45 (1981): 233–256; see also David Lewis, "Causal Explanation," *Philosophical Papers*, vol. 2.

the earlier one, and, therefore, the two explanations are not independent. This, too, reflects realist thinking: two explanations are thought to be nonindependent because the explanatory relations represented by them are not independent.

(5) c_1 is part of c_2.[26] The explanations, then, are not independent; nor can they both be complete.

(6) c_1 and c_2 are independent, each a sufficient cause of e. This, then, is a standard case of "causal overdetermination." Do we in this case have a counterexample to the explanatory exclusion principle? Why aren't both explanations, "e happened because c_1 caused it" and "e happened because c_2 caused it," sufficient and independent explanations? This is an interesting case from the point of view of both explanatory exclusion and the question of explanatory realism versus irrealism, and we shall discuss this in some detail.

Hempel has called cases like this "explanatory overdetermination":[27] Suppose that a copper rod is heated while simultaneously being subjected to longitudinal stress. As a result, its length increases. Two deductive-nomological (hereafter "DN") arguments can be formulated: the first would invoke the lawlike premise that copper rods lengthen when they are heated, and the 'initial condition' that this particular copper rod was heated on this occasion; the second would appeal to the law stating that copper rods increase in length when subjected to longitudinal stress, and the initial condition that this copper rod was subjected to that kind of stress. The two arguments share the same conclusion, the statement that the rod's length increased on this occasion. According to the standard DN account of explanation, therefore, each argument counts as an explanation.

It is not surprising that Hempel rejects the view that these DN arguments are not "complete" as explanations. He writes:

> It might be objected that—even granting the truth of all the premises—both accounts are unacceptable since they are "incomplete": each neglects one of the two factors that contributed to the lengthening. In appraising the force of this objection it is again important to be clear about just what is to be explained. If as in our example, this is simply the fact that *Lr*, i.e., that *r* lengthened, or that there was *some*

[26] Karl Pfeifer brought this case to my attention.

[27] Carl G. Hempel, *Aspects of Scientific Explanation* (New York, 1965), pp. 418–420.

> increase in the length of *r*, then, I think, either of the two arguments conclusively does *that*, and the charge of incompleteness is groundless.[28]

Here he seems simply to affirm that, as an explanation of why the rod lengthened, "each of the two arguments conclusively does *that*." But why does he say this? The use of the term "conclusively" suggests that he was moved by the consideration that each DN argument provides a premise-set that is *deductively conclusive* for the truth of the explanandum statement. This is not surprising. For, fundamental to the DN conception of explanation is the idea that explanations are *inferences* or *arguments* of a certain form. Given this assumption, a natural sense of "completeness" or "sufficiency" emerges for explanations: when an argument has the correct DN form, it is *complete* and *sufficient*. Hempel writes:

> I think it is important and illuminating to distinguish such partial explanations . . . from what might be called *deductively complete explanations,* i.e., those in which the explanandum as stated is logically implied by the explanans; for the latter do, whereas the former do not, account for the explanandum phenomenon in the specificity with which the explanandum sentence describes it. An explanation that conforms to the D-N model is, therefore, automatically complete in this sense; and a partial explanation as we have characterized it always falls short of being a D-N explanation.[29]

As explanation is conceived under the DN model, there is nothing one can do to a DN argument to improve it in regard to its "completeness" as an explanation. One may be able to make it deeper, more perspicuous, more systematic, and so on; but what could one possibly do to make it "more complete"? The DN conception of explanation does not seem to leave room for any other sense of explanatory completeness than deductive conclusiveness.

These considerations suggest that a preoccupation with the deductive or inferential character of explanation leads to a form of explanatory irrealism ("explanatory internalism," perhaps, is more appropriate), and this is certainly what we see in Hempel. This internalist tendency is evident also in

[28] Ibid. 418–420.
[29] Ibid. 416–417.

Hempel's well-known emphasis on the predictive character of explanation, and in one of his two conditions of adequacy on explanations, that is, the requirement of "explanatory relevance" to the effect that "explanatory information adduced affords good grounds for believing that the phenomenon to be explained did, or does, indeed occur."[30] Hempel's idea that explanations are arguments, his condition of "explanatory relevance," and his emphasis on the predictive aspect of explanations go hand in hand: they all point to explanatory irrealism—at least, point away from explanatory realism with the causal relation serving as the explanatory relation.

Hempel's primary focus in analyzing the structure of explanation is on the logical and conceptual characteristics of statements making up an explanation (the "internal" properties, as I have called them), not on the events or other entities these statements describe and their interrelations. In fact, we get from Hempel a precise and elaborately constructed definition of what an explanation is, but only a very intuitive and unanalyzed idea of what it is that a given explanation is an explanation *of*.[31] Hempel's treatment of causal explanation and causation is also symptomatic of this attitude:[32] the idea of a DN argument, an essentially internal notion, is primary in the characterization of explanations, and the idea of causal explanation falls out of this characterization as a not-so-clearly-defined special case. Hempel evidently does not regard the concept of causal explanation, or that of causation, as at all crucial to a theory of explanation; his discussion of causal explanation often comes across as a concession to the popular practice of referring to causes and causal explanations, not something that he sees as essential to the development of his theory. From such an internalist perspective, it is entirely natural that each of the two DN arguments about the expanding copper is regarded as complete and sufficient in itself as an explanation.

What does explanatory realism say about the expanding copper rod? If the heating and the stress are each an independent sufficient cause of the

[30] *Philosophy of Natural Science* (Englewood Cliffs, NJ, 1966), pp. 47–49. See also his *Aspects of Scientific Explanation*, pp. 367–368. The other adequacy condition is the unexceptionable requirement of "testability," to the effect that the explanatory premises must be capable of empirical test in a broad sense.

[31] For further elaboration of this point and some suggestions, see my "Events and Their Descriptions: Some Considerations," in *Essays in Honor of Carl G. Hempel*, ed. Nicholas Rescher et al. (Dordrecht, 1969).

[32] Hempel, *Aspects of Scientific Explanation*, pp. 347–354.

rod's lengthening, we have a standard case of causal overdetermination. Moreover, if, as explanatory realism seems to suggest, explanatory completeness is to be understood in terms of sufficient cause, it follows that in the present case we have two independent and complete explanations. Thus, explanatory realism seems to yield the same result as Hempel's irrealism: both seem to contradict the explanatory exclusion principle.

The explanatory realist who wants to save explanatory exclusion might deny that the rising temperature and the stress were each a sufficient cause of the event to be explained, and deny, more generally, that genuine instances of causal overdetermination exist. Peter Unger has claimed that each event has a single unique cause (at most),[33] and if this is right, then not both the heating and the stress can be a cause of the lengthening. Therefore, there could be at most one causal explanation here. But Unger's thesis is a radical one, too strong to be plausible: he construes it to entail the denial of transitivity of causation, and hence the impossibility of causal chains with more than two links. And, his arguments rely exclusively on a certain kind of linguistic evidence whose point I find difficult to evaluate.

Martin Bunzl, too, has argued that there are no genuine cases of causal overdetermination.[34] His basic point is that the usual examples, when closely scrutinized, turn out to be either cases of causal preemption or of joint cause. That is, one of the two alleged overdetermining causes preempts the other (by "getting there first") so that the second, in fact, is not a cause of the effect in question, or else the two causes together make up a single sufficient cause, neither of them alone being sufficient. I think Bunzl's arguments, on the whole, are plausible, though not conclusive.[35] Thus, when applied to the case of the copper rod, his analysis would probably give this diagnosis: the particular lengthening that took place was caused by the single joint cause made up of the heating and the stress. Neither of the

[33] "The Uniqueness in Causation," *American Philosophical Quarterly* 14 (1977): 177–188.

[34] "Causal Overdetermination," *Journal of Philosophy* 76 (1979): 134–150.

[35] Bunzl, however, says that his considerations depend essentially on a certain view of the nature and individuation of events associated with Donald Davidson, and that they are ineffective if we assume the sort of view of events that I myself have advocated, namely, one that takes events as property-exemplifications. See Bunzl, "Causal Overdetermination," p. 150. However, I am not convinced of this; I think Bunzl may have been misled by just the kind of consideration that lead Hempel to believe in explanatory overdetermination. It is interesting to note that Bunzl accepts explanatory overdetermination in Hempel's sense (p. 145). On causal overdetermination, see also Louis E. Loeb, "Causal Theories and Causal Overdetermination," *Journal of Philosophy* 71 (1974): 525–544.

two events was, in itself, a sufficient cause of it. Thus, a complete explanation of the lengthening must refer to both the heating and the stress as a single sufficient cause.

We must set aside the question whether genuine instances of causal overdetermination exist. What is of interest to us here is that under explanatory realism, the causal relation can be made to do some real work, in characterizing and constraining explanations. As we saw, the association between causation and explanation, underwritten by explanatory realism, yields a principle of individuation for explanations and a notion of "complete explanation," both essential to interpreting the principle of explanatory exclusion. We also saw that if causal overdetermination is not possible, that takes away one potential case of explanatory overdetermination. It seems to me that we are inclined to take these considerations involving causation as both natural and relevant in discussing the nature of explanation. What accounts for this inclination, I think, is our tacit acceptance of explanatory realism: for a causal explanation to hold, the explaining event must be a cause of the event explained. Given this connection between causal explanations and causal relations, we are able to use facts about the latter to say something about the former.

To return briefly to the matter of explanatory exclusion: if our considerations are generally right (especially in the treatment of the six cases in which two causal explanations are offered for one event), explanatory realism is seen to provide a sense, as well as support, for the explanatory exclusion principle—except, perhaps, in the case of causal overdetermination, which we set aside without a clear-cut resolution.[36] I believe it is more difficult, though not impossible, to interpret and argue for explanatory exclusion if by embracing explanatory irrealism we lose the causal handle on explanation.[37]

[36] From the point of view of explanatory exclusion, causal overdetermination is not crucial; the exclusion principle has content of sufficient interest even if causal overdetermination is simply exempted.

[37] I think explanatory exclusion can hold under explanatory irrealism as well; however, unlike explanatory realism, irrealism does not, I think, provide a positive basis for explanatory exclusion. For some considerations favoring explanatory exclusion that are not based on explanatory realism, see my "Mechanism, Purpose, and Explanatory Exclusion."