

Causation

DALE JACQUETTE, *Universität Bern, Institut für Philosophie, Lehrstuhl für theoretische Philosophie, Unitobler, Länggassstrasse 49a, CH-3000 Bern 9, Switzerland.*

E-mail: dale.jacquette@philo.unibe.ch

Abstract

The interconnectedness of all events in the causal matrix suggests that the so-called informal fallacy of *post hoc, ergo propter hoc* reasoning is not deductively invalid. It may be preferable then to consider *post hoc, propter hoc* as violating a pragmatic concept of causal relevance. A leading pragmatic account of relevance defines it as the caused closure of a cognitive agenda, to which it seems necessary to add that such closure be relevantly caused, and not by such contingencies as the inquirer's death or incapacity, magic philosopher's pills or the like. This in turn involves the concept of relevant causation in a vicious circularity. The implication may then be that relevant causation, like time and space as pure forms of intuition in Kant's Transcendental Aesthetic, cannot be defined or reductively characterized in terms of more primitive concepts, let alone derived empirically or pragmatically from the contents of experience or purpose of an action. Causation and causal relevance on this proposal constitute instead part of the mind's innate quasi-metaphysical bedrock of scientific explanation.

Keywords: Agenda relevance, causal relevance, causation, Dov Gabbay, David Hume, Immanuel Kant, *post hoc, propter hoc* fallacy, pragmatism, relevance, relevant causation, John Woods.

1 The causal nexus

As David Hume taught in *A Treatise of Human Nature*, explaining the experiential origin of the concept of causation, any kind or category of event in principle can cause an event of any other kind or category [17, Book I, Part III, Sections II, VI and XIV]. Causation is logically contingent, but causal connections can further be shown also to be considerably weaker than other logically contingent but still logically related propositional relations, such as supervenience in the sense of unqualified ontic dependence.¹

To appreciate Hume's point about causal modality, we might begin by reflecting on the *post hoc, propter hoc* fallacy. This supposedly invalid form of reasoning is more properly classified, if a fallacy at all, as an informal extralogical fallacy of explanatory relevance, rather than a colloquially expressed inference cloaking a formal logical deductive invalidity. For, in a sense, there is nothing logically fallacious about *post hoc, propter hoc* reasoning at all. The interconnectedness of all spatio-temporal events throughout the world's vast causal matrix is already implied by Isaac Newton's discovery that gravity is universal. If every physical entity anywhere in the universe exerts a gravitational attraction on every other physical entity, however weak and for explanatory purposes practically inconsequential, if, secondly, gravitational forces have causal efficacy, as they surely do, and, finally,

¹The most straightforward argument for the logical contingency of causal connections is that there is no syntactical contradiction or tautology in the complete and correct description of any proposition asserting the existence of an actual cause and another denying the existence of the effect. Hume, George Berkeley and empiricists generally have relied on one version or another of this line of reasoning. See, *inter alia* [2, 10]. That there might another, extralogical, necessity obtaining between cause and effect is untouched by the logical contingency of causal connections.

if causation is a transitive relation, then there is a causal connection between any past event, however remote in time or space, and any later event for which we seek to establish a causal explanation.²

Thus, when someone reasons concerning any two events x and y that y happened at some time after event x and therefore that y happened because of event x , no matter how otherwise unrelated x and y may be, the inference cannot be dismissed as deductively logically invalid. We assume throughout the truth of such principles, fundamental to chaos theory and Newtonian physics, that link together all events in the causal matrix, as the causal efficacy of universal gravity. It will in fact be true in that case that any temporally preceding event will be a cause of any later occurring event, which is to say that it is impossible, contrary to the import of the supposed *post hoc, propter hoc* fallacy, that we can ever possibly begin with the true assumption that a given event x has occurred in time *before* the occurrence of another event y , and infer from that assumption to a false conclusion that therefore event y occurred *because* of the prior occurrence of event x . Whatever other status the *post hoc, propter hoc* inference schema might have, it cannot be rejected as a logical or deductive fallacy in light of the far-reaching causal nexus by which all events are contingently linked together in spacetime.³

2 Counterfactuals in the analysis of causation

A possible response to the argument against *post hoc, propter hoc* as a logical fallacy is to say that the complete analysis of causal connections must introduce considerations of lawlike but still logically contingent counterfactual dependence. If y is caused by x , then had x not occurred, y would also not have occurred.⁴

Counterfactual dependence is no doubt a part of the intuitive conception of causation that logicians and philosophers have sought to explicate. In doing so, they have often made use of the powerful devices of set theoretical analysis of branching states of affairs across alternative logically possible worlds. However, we might find grounds for scepticism about the accuracy of such modellings,

²It is maintained here, perhaps somewhat surprisingly, that it is not logically possible for an event E_1 to occur before another event E_2 but for E_2 not to occur because of E_1 . The temptation to posit a deductive invalidity here is due in part perhaps to the fact that both the proposition stating that E_1 occurs before E_2 and the proposition that E_2 occurs because of E_1 are logically contingent. It may appear inescapable, then, that the fact that E_1 occurs before E_2 is no logical guarantee that therefore E_2 occurs because of E_1 . Formulated in just this way, there is indeed much to recommend the inference as logically fallacious, since in general no two logically independent contingent propositions can be such that if the one is true then the other must logically be true. What is missing from the picture, however, is the assumption we emphasize that every event of the past is a causal factor in the occurrence of every temporally subsequent event. Combining this assumption with the proposition that E_1 occurs prior to E_2 now makes it logically impossible for event E_2 not to occur because of E_1 . The inference in the opposite direction has also been considered by Rosenberg [25].

³I interpret Hume's telling phrase 'the cement of the universe', including Causation as one of three 'links' or 'principles of association', along with Resemblance and Contiguity, as binding every part of the universe together into one great unity. Hume writes, in 'An Abstract of a Book Lately Published, entitled *A Treatise of Human Nature &c.*' [1740] [17, Section 35]: 'It will be easy to conceive of what vast consequence these principles must be in the science of human nature, if we consider that, so far as regards the mind, these are the only links that bind the parts of the universe together, or connect us with any person or object exterior to ourselves. For as it is by means of thought only that anything operates upon our passions, and as these are the only ties of our thoughts, they are really to us the cement of the universe, and all the operations of the mind must, in a great measure, depend on them.' The idea undoubtedly originates with Isaac Newton's theory of universal gravitation. For historical background, see [19].

⁴Mackie [22] is sometimes cited as offering the most highly developed counterfactual analysis of causal connections, although Mackie himself considers the account to be conceptually incomplete. Other counterfactual treatments of causation in a substantial literature include the studies [3, 7, 8, 21, 24]. A useful collection of essays on the topic, reflecting especially on Lewis's 1973 paper, is found in [5].

when as always we are possessed at best of only partial information about the real contribution that apparently causally irrelevant past events might exert on later events. It is thus an established part of conventional wisdom in talk about causal relations to hold that the assassination of Abraham Lincoln in 1865 is causally irrelevant to a match igniting when struck in 2009. The expectation is that if Lincoln had not been assassinated then there seems to be no reason to think that the match would not catch fire. Perhaps the counterfactual proposition is true. But how do we *know* this? It is one thing to construct a formal model in which the non-occurrence of Lincoln's assassination takes place in a logically possible world accessible by an appropriate cocktail of model set theoretical relations to a logically possible world in which at a later time the match lights when struck. It is quite another thing to know with respect to events in the actual world that, if a seemingly causally irrelevant past event had not occurred then a later event would definitely not have occurred.

Common sense also supports the view. The counterfactual is nevertheless unmanageable insofar as it pretends to describe the real causal dependencies governing the actual world in which the assassination and the lighting of the match take place. How do we know that if Lincoln had not been assassinated as happened in fact that the American Civil War might not have resurged, resulting in a destruction of the prepubescent ancestors of the person who in the actual world actually strikes the match? Or that in unanticipated ways the non-assassination of Lincoln might not have led to the development of new technologies such that in 2009 there simply are no matches? I submit that when we bandy about these counterfactuals and claim to know what would or would not happen in the future following the non-occurrence of some actual event in the past, we are merely playing with words lacking any verifiable factual scientific or logically or philosophically sound metaphysical foundation.

Undoubtedly an important distinction needs to be invoked here in any case between causation and causal explanation. We might reasonably say that the match would not have ignited if the Big Bang had not occurred. Yet it would be difficult to maintain, given the practical purposes under time constraints for explanations and their application, that the match ignites because of the Big Bang and all the great chain of causally related events intervening from then until the moment when the match is struck.

3 Causal token and type events

Another consideration that weighs in favour of disregarding *post hoc, propter hoc* as a deductive fallacy is the fact that identity conditions for events might need to be understood holistically in terms of their position and interrelation with other past, present and future events in a grand encompassing network of spatio-temporal occurrences.

What makes a particular event of a man's being killed by an assassin or a match being struck, the thing in question for which a causal explanation is sought, might well be bound up with its relations to all the events of the past, and in particular with the fact that one event occurs temporally after the other. Even on Leibnizian grounds, the identity conditions for the specific event in which a given match catches light in such and such a place and such and such a time might need ultimately to make reference to the fact that the event of the match igniting occurs at precisely the time it does after all previous events that happen to have transpired throughout the wide universe, including Lincoln's assassination. If this is true, then thinking away the occurrence of Lincoln's assassination might leave us entirely bereft of the specific event whose cause we would fain understand. Identity conditions are highly demanding and, when properly understood, imaginatively tamperproof.

A promising reply to these objections is to enforce a sharp distinction between token–token versus type–type causal explanations. It might well be true that we cannot advance token–token causal explanations that are supported by counterfactuals bearing significant distinctions between causally relevant and causally irrelevant past events, for which, as a consequence, *post hoc, propter hoc* is after all a deductive logical fallacy. Type–type causal explanations might offer a rather different picture, and one to which it might also be possible to link causal accounts of specific token events in a modified version of the Hempel–Oppenheim nomological–deductive model of scientific explanation. Although we might not be able in good conscience to eliminate Lincoln’s assassination as a cause or causal factor in a match’s igniting 144 years later, we might be able, first, to exclude past presidential assassinations as an event-type causal factor in present day event types in which matches are made to light. After all, no one who wanted a match to burn would sanely consider as part of the process making sure that an American President had been assassinated. Dry matches, proper striking surface, angle and force of striking, exact chemical preparation of the match tip, presence of oxygen, yes, those things are on the list; assassinate the President, or make sure that an American President had been assassinated at least a century ago, in order to make the match light, as a general thing, is not. We might, therefore, reasonably conclude that the event type of a presidential assassination is causally irrelevant to the event type of making a match flare. Then we might set about explaining the cause of a particular match’s lighting as a historically situated specific token event by referring to the event-type correlations in which meaningful discriminations are made in roughly the common sense way between causally relevant and causally irrelevant factors. We take the opportunity accordingly to distinguish between token- and type-causation, to indicate the difference just explained involving causal explanations between event tokens versus event types.⁵

The causal explanation model recommended by these considerations then has the following form:

- (1) Type B events are type-caused by type A events.
 - (2) Token event y is a type B event.
-
- (3) Probably, token event y is type-caused by a type A event.

Notice that the above model does not readily accommodate, if it accommodates at all, causal explanations of token events in terms of other token events. Thus, we cannot continue the proposed chain of reasoning in this way:

- (4) Token event x is an type A event.
-
- (5) Probably, token event y is token-caused by token event x .

There will be unlimitedly many scores of instances of most event types, and there is no guarantee that the token event consisting of the presence of oxygen in dynastic Egypt is a cause in the token event of a match’s bursting into flame in the Arctic circle in 2009.

Evidently, where we indicate the time in some choice of units t , $n \geq 1$ at which an event E_N occurs, it would be patently deductively invalid to infer:

- (1) $\text{Occurs}^{t-n}(E_1) \wedge \text{Occurs}^t(E_2)$
-
- (2) $\text{Causes}(E_1, E_2)$ INVALID!

⁵On applications of the token-type distinction in recent causation theory, see especially [14].

The argument is nevertheless deductively valid if we regard the inference as enthymematic for the properly expanded argument, based on an understanding of the causal role of total background conditions to any subsequent effect:

(1) $\text{Occurs}^{t-n}(E_1) \wedge \text{Occurs}^t(E_2)$

[1a. $\forall x[[\text{Occurs}^{t-i}(x) \wedge \forall y[\text{Occurs}^t(y)]] \rightarrow \text{Causes}(x,y)]$]

(2) $\text{Causes}(E_1, E_2)$

VALID

Clearly, the soundness of the second inference form depends entirely on interpolated assumption [1a]. The justification for the proposition once again is the interconnectedness of the causal nexus and our inability to negotiate unmanageable counterfactuals to determine confidently and with strong epistemic assurance that any given past event is definitely not causally involved in the production of any later events, as the effect of universal gravitation is meant to illustrate. The expanded inference above is not made trivially valid, as long as we have good philosophical and scientific grounds for asserting the truth of [1a].

It may appear that Julius Caesar entertaining a thought about what to have for dinner as he is crossing the Rubicon could not causally contribute to the occurrence of an earthquake hundreds of years later. Such psychological occurrences are nevertheless physical events involving the movement of physical substances in a thinking subject's neurophysiology, and as such they cause an unknown and unlimited number of subsequent events, which, through the so-called multiplier effect, might for all we know later be causally involved in the occurrence of an earthquake. In any case, such physical events as a subject's thinking something at an earlier time are undoubtedly part of the total background conditions that collectively are causally responsible for whatever happens holistically at subsequent moments within the vastly complexly interconnected causal nexus. The inference in *post hoc, propter hoc* reasoning therefore need not be regarded as deductively invalid, provided that we find it reasonable to suppose that the occurrence of any event is actually caused by the total background of prior physical states in the universe as a whole.

4 Gabbay–Woods and the pragmatics of agenda relevance

The better tactic by far would therefore seem to shift focus entirely away from *post hoc, propter hoc* as any kind of *deductive* fallacy of *reasoning* or *logical inference* to its reinterpretation as an informal extralogical fallacy of *explanation*. It is not that we are ever in a position correctly to judge that it is logically possible for a past event to have occurred and a later event not to have occurred. The question is rather one of explanatory utility.

According to this way of thinking, we are never in a position correctly to judge that an instance of *post hoc, propter hoc* transition from assumptions to conclusion is somehow logically invalid. For, we are never in a position even on the counterfactual model correctly to judge that it is logically possible for a past event to have occurred and a later event not to have occurred. Rather, we might conclude instead that *post hoc, propter hoc* is a fallacy that sometimes besets our efforts to provide a good explanation of cause and effect in a more pragmatic sense. We seem in that case to have distinctively lowered our expectations about the sense in which the cause of an event is to be identified than that by which the meaning of cause, however it is finally spelled out, is regarded in a more strict, scientific and metaphysical sense, according to which any prior event can be truthfully regarded as at least part of the cause or indispensable causal background for a later occurring event.

If, now, we choose to shift focus from deductive inference to the pragmatic requirements of explanation, where might we turn for guidance? The most complete and thoroughly developed pragmatic theory of relevance to date is that of Gabbay and Woods. Their book, *Agenda Relevance: A Study in Formal Pragmatics*, Volume 1 in the series, *A Practical Logic of Cognitive Systems*, proposes a three-part analysis of the concept of relevance. Relevance, according to Gabbay and Woods's pragmatic theory, is itself a *causal* relation defined over ordered triples $\langle I, X, A \rangle$ of *information*, *cognitive agents* and their *agendas* [13, p. 74].

The idea of defining relevance in relation to an agenda offers an interesting way of unifying a large number of pretheoretical intuitions about what makes something relevant. An agent has something to do, a purpose or goal to accomplish. Gabbay and Woods distinguish between such related concepts as plans and strategies, and analyse agendas more precisely as artefacts of theory, abstractions from 'plans, intentions, strategies, tactics, functions, designs, programs, scripts, tasks, undertakings, conventions and dispositions' [13, pp. 204–205]. Roughly speaking, relevance in Gabbay and Woods's sense is a matter of how information bears causally on the closure of a cognitive agent's cognitive or other practical agenda via belief revision. Gabbay and Woods distinguish agendas from the related concepts of goals, aims or purposes. However, if we think crudely for the moment of an agenda as among an agent's purposes, then a specific item of information is judged relevant relative to an agenda just in case it causally contributes to accomplishing or 'closing' the agenda.

The analysis of agenda relevance provides a framework that combines a variety of independently interesting factors, each possessing a range of explanatorily exploitable dimensions. The fact, for example, that the theory is supposed to account for the causal contribution of information to an agent's agenda allows for external criteria of success in relevance judgements. We should in principle be able to determine whether or not a given item of information is or is not relevant to an agent once we identify a specific agenda and make the appropriate discovery of the causal impact specific items of information are likely to have in completing the agenda. Thus, if an agent wants to know whether the sun will shine during a planned picnic, and the movement of a barometer's needle indicates a sudden falling of atmospheric pressure that generally presages rain, then that information can positively contribute causally in a confirmable way to closing the agent's cognitive agenda by satisfying the desire to know what sort of weather to expect. The agenda is then closed in the sense that the agent has thereby been caused not to continue inquiring, but to consider the matter resolved and to move along to the problems posed by another appropriately prioritized agenda in the agent's queue.

Gabbay and Woods reasonably assume that practical agents must always try to make economical use of scant cognitive resources in order to pursue their purposes and prosecute their agendas with an eye to pragmatic satisfaction and success. We have only limited time, limited access to information and urgent tasks to perform within these constraints. It is through these limitations that the problem of relevance arises in the first place, by requiring that we have a sound way of screening out information that is not pertinent to achieving our goals or fulfilling our agendas, and focusing to whatever extent possible only on useful information that helps us to accomplish whatever we need and want to do. The challenge in its essentials is the same for us in all our cognitive activity as the combination of what the authors identify as the *knowledge-organization problem*, *frame problem* and *relevance problem* in artificial intelligence, as discussed in contemporary cognitive science [13, 109–110].⁶ The task itself under practical limitations calls upon epistemic agents subcognitively to exercise discretion in excluding as noise information that is irrelevant to whatever purpose is at hand.

⁶Among the vast literature on the frame problem, see also [6, 11, 32].

The pretheoretical concept of relevance on which Gabbay and Woods's formalization is grounded distinguishes between propositional relevance, legal relevance and topical relevance, and requires that the formal analysis of relevance do justice to all of these specialized applications at a higher level of generality. Altogether, Gabbay and Woods introduce 10 adequacy conditions for a correct theory of agenda relevance. The analysis is supposed to: (i) avoid trivially entailing either that nothing is relevant to anything or that everything is relevant to everything; (ii) entail that relevance is context-sensitive; (iii) admit of comparative degrees of relevance; (iv) provide for a relation of negative relevance in addition to mere irrelevance; (v) help elucidate the fallacies of relevance; (vi) explain the distinction and rivalry between classical and so-called relevance logics; (vii) contribute to the theory of belief revision; (viii) contribute as well to the suggestion that relevance is intrinsically a dialogical concept; (ix) provide a general account of all legitimately designated kinds of relevance; while (x) accommodating worthwhile insights from alternative theories of relevance [13, pp. 155–156].

Gabbay and Woods hold their theory to a well-articulated vision of the needs and reasonable expectations of logical theory, judging them against such parameters as Grice's condition to preserve relevance in the common sense understanding of the concept, further interpreted by the so-called SW-relevance theory of Sperber and Wilson [28, 29].⁷ The latter account provides a widely accepted benchmark with which Gabby and Woods sometimes agree and sometimes disagree. They additionally require their theory to conform to finely nuanced grammatical distinctions to which they attach philosophical significance, such as that, most notably, between *relevance to* certain belief-adjusting factors of an agent, and *relevance for* the agent. These conditions in turn sometimes but do not always coincide, and the belief adjusting factors in question are naturally subdivided into what the authors colourfully describe as *Seers of Trouble Coming* and *Putters of Things Right*. The idea is that in both of the above types of belief adjusting a factor could be relevant to an agent's agenda in several ways without the agent recognizing it as such, taking it into account or putting it to use. The desiderata of agenda relevance in turn support a concept of *meta-agendas*, whereby agents can have a hierarchy of agendas in which higher order agendas impose demands that lower order agendas be completed in particular ways, say, with economy and efficiency or even with a certain aesthetic finesse [13, pp. 225–230].

5 Circularity objections to Gabbay and Woods's analysis

The Gabbay–Woods conceptual–logical pragmatic analysis of relevance is no sooner articulated than we encounter an apparently insurmountable obstacle to its applicability in understanding causal relevance and relevant causation. The trouble is that Gabbay and Woods treat relevance itself as a causal relation.

'Relevance is a causal relation', they explain. 'Among other things, it causes or helps cause changes of mind' [13, p. 163]. This commitment makes the theory of agenda relevance dynamic, verifiable and externalist in attractive ways, but it also appears to incur a number of unfortunate theoretical liabilities. For one thing, it immediately implies that a relevant factor must actually cause or help to cause an agent's belief revision, which presumably can occur only when a factor is relevant *for* an agent rather than merely *to* an agent's agenda. Even greater difficulties loom. It appears, for example, that some factors could be relevant to an agent's agenda without ever actually causing or helping to cause a revision of belief and without ever actually causing or helping to cause an agent's agenda to close.

⁷Among the vast literature on the frame problem, see also [27].

Suppose that it is my cognitive agenda to discover whether or not it is raining. The fact that the sun is shining in the relevant geographical region would seem to be eminently relevant to my agenda. It could also be relevant for me as an epistemic agent if I know about it and if my acquaintance with the facts were consequently to revise my belief from a state of curiosity, suspended belief or unknowing to one of having in hand the answer to my query. However, if I do not become familiar with the fact that the sun is shining, then my agenda is never actually closed. Nor does the mere fact that the sun is shining by itself actually cause or help to cause any revision in my belief state or subsequent practical course of action. It would do so, of course, and it could potentially do so, *if* I became aware of the fact; but we are imagining that it does not actually do so, despite being highly relevant to my agenda. Thus, another thinker from a third-person perspective could reasonably judge that such information was relevant to my agenda, while recognizing that there is no actual but at most only a subjunctive counterfactual causal impact on my belief states. The problem seems above all to cast doubt on the plausibility of Gabbay and Woods's distinction between *relevance to* and *relevance for*. Relevance to an agent's agenda appears to involve the sort of manifest causal efficacy that Gabbay and Woods require only when it is relevant for the agent; otherwise hypothetically no belief revision is actually caused. Gabbay and Woods do additionally consider the concept of *relevance potential* [13, pp. 184–188]. This category is nevertheless reserved specifically for information an agent has taken in subliminally or subconsciously, but that at a given time has not yet actually caused any belief revision. What Gabby and Woods do not seem to consider is the potential for information that has not actually touched an agent and has not been cognitively assimilated, but that might still be said intuitively to be relevant and not merely potentially relevant to the agent's agendas, in the sense that if the agent were to take in the information then it could or would probably bring about a revision of the agent's beliefs that could or would probably close or contribute to closing a particular cognitive agenda.

It might be thought that it would be a relatively easy matter to amend the Gabby–Woods analysis to countenance the potential causal force of information in causing specific types of belief revision. Such a manoeuvre, unfortunately, as reasonable as it may at first appear, soon leads to another difficulty. Belief revision might be caused by an agent's assimilation of unlimitedly many different kinds of information in unlimitedly different kinds of ways. If causation, as Hume thought in uncovering the origin of our idea of cause and effect, is just the regular concurrence of temporally ordered spatially neighbouring events, then there is no reason not to suppose that my coming to know that 3 is a prime number might *cause* me to change my belief that it is raining, and hence to close one of my cognitive agendas. When we consider the wide range of possibilities for agenda closure, then information of almost any type might cause changes in beliefs of any type, in the same way that hypnosis, a philosopher's pill or a sharp rap on the head might in principle, as a logical possibility, induce a subject to believe anything we please. If I have grasped the conceptual analysis of agenda relevance that Gabby and Woods endorse, then, in the above scenario, although the conclusion seems preposterous, the information that 3 is a prime number would be relevant to my cognitive agenda of wanting to know whether or not it is raining. What one might otherwise perhaps like to say here is that, although coming to learn that 3 is a prime number might cause me to revise my belief about whether or not it is raining, this new information about 3's being prime does not *relevantly cause* my change of belief, and that consequently it does not cause my belief to be modified or revised in a *relevant way*. If that is the natural reaction to such a situation, however, then Gabbay and Woods have evidently not yet pursued the analysis to its ultimate conclusion, but have offered an analysis that presupposes a prior understanding of the concept of agenda relevance or of some type of casual relevance, or perhaps of generic relevance as such, that threatens to make their analysis incomplete if not viciously circular.

Such internal theoretical circularity is bad enough. Gabbay and Woods's pragmatic analysis of agenda relevance is regrettably embroiled in a deeper circularity when applied in particular to the concept of causal relevance and relevant causation. If we hope to understand what is meant by causal relevance and relevant causation, then we can hardly hope to find comfort in a theory that presupposes exactly those concepts. If, as suggested here, it is not sufficient merely for a certain factor to cause an epistemic agent's cognitive agenda to close, but for it to do so in a causally relevant way to be explained by invoking the concept of relevant causation, then an especially hazardous circularity threatens the entire enterprise, both of understanding causal relevance from the kind of pragmatic perspective represented by Gabbay and Woods's analysis of agenda relevance, and ramifying hence to menace the understanding of agenda closure in its most general terms.⁸

6 Causation, causal relevance and relevant causation as primitive

To recap the argument in a nutshell, we began by remarking that to fully understand causation and causal relations or connections, we need to address the sense in which *post hoc, propter hoc* is a fallacy. If, as it seems, it is not deductively invalid to infer that any past event has caused any future event, given the interconnectedness of the spatio-temporal causal nexus of total background conditions to the occurrence of any later event, then we must either tolerate such reasoning as correct, which in fact we do not, or we must cast about for another category of objection and another category of fallacy to this type of inference from effect to presumptive cause.

The most natural way to think about *post hoc, propter hoc*, we have suggested, is to consider the fallacy not as a deductive invalidity or logical fallacy in the narrow sense, but rather as a failure to satisfy the pragmatic requirements of good causal explanation. When, in turn, we look to the analysis of what constitutes good causal explanation, we are driven at once to recognize the indispensability of causal relevance and hence of relevant causation. Indeed, we might rightly insist that to speak of causation at all is necessarily to speak of relevant causation. We do not consider something a *cause* in the first place unless we believe that its occurrence is causally relevant to the production of a certain effect. The next step logically, then, is to investigate the concepts of causal relevance and relevant causation, as particular instances of the more general concept of relevance. Where these two lines of inquiry intersect, we are led to consider more specifically a pragmatic analysis of the concept of relevance.

The details of Gabbay–Woods's pragmatic analysis of relevance on critical scrutiny unfortunately reveal circularities of two distinct but related types. The arc of the first circle begins when we try to define causal relevance and relevant causation in terms of the causal closure of a cognitive agenda, acknowledging that merely causing a cognitive agenda to close in any old logically possible way is not good enough to provide a pragmatic explanation of the relevance of whatever it is that causes the agenda to close. The circle is completed when we further recognize that in order to secure a satisfactory pragmatic explanation of why the agenda closes upon the attainment of particular items of information, we must appeal to a prior understanding of an agenda's closure being relevantly caused. The second circle is even tighter, though it leans heavily on the first. It is set in motion when we consider that to properly understand the concept of causal relevance and relevant causation, we may need to invoke a more general pragmatic theory of relevance, and is completed when we reflect that any pragmatic theory of relevance like Gabbay and Woods's, for reasons of its entanglement in the first circularity presupposes rather than helps us to independently define the twin concepts of causal relevance and relevant causation.

⁸ See Jacquette [18] from which I have extracted certain parts of this critique.

What, philosophically, then, are the options that remain in light of these two circularity objections? Assuming that the circularity arguments cannot be overturned, what attitude should we adopt toward causation, relevance, causal relevance, relevant causation and pragmatic accounts of agenda relevance generally and in Gabbay and Woods's case in particular? We can see that one otherwise tempting way out of the difficulty is altogether hopeless. It might appear that we could escape the circularities by blaming the problem on Gabbay and Woods's pragmatic analysis of agenda relevance. To zero in on what seems to be the main source of circularity, why in the first place should we agree with Gabbay and Woods that agenda closure is *caused* by agenda-relevant factors? The answer seems to be that we have no choice in the matter, and that Gabbay and Woods have put their collective finger on something that is absolutely essential to agenda relevance. It is natural and unavoidable, if we want to do justice to our reflective pre-analytic intuitions about the concept of relevance, to think of agenda relevance as the need for information that is literally caused to be closed or satisfied when the information is acquired, as testified by the agent's subsequent actions and the projected success in the agent's pursuit of practical purposes.

To whatever extent we are attracted to other features of Gabbay and Woods's analysis of agenda relevance, if there is any aspect of their account that cannot lightly be pruned away, it is surely their recognition that the right information causes a cognitive agenda to close. They are undoubtedly correct to maintain that there is a specifically causal connection between an agent arriving at information relevant to a purposeful activity and the cessation of further cognitive efforts in the respective purposive quest. When we learn what we need to know in order to do something, we are *caused* to call a halt to further search. We go badly off-target if we turn away from Gabbay and Woods's analysis of agenda relevance in favour of another theory that does not pay heed to the causal efficacy of information in bringing closure to cognitive agendas as a central element in what makes information relevant to an epistemic agent's pragmatic needs and goals.

If we have no option but to admit that relevance must be defined as the caused, and, even more trenchantly, the relevantly caused closure of a cognitive agenda, then we may also have no alternative but to acknowledge that relevant causation, which, on final analysis, is simply to say causation, is an inherently circular concept, one that is not to be grounded foundationally in more primitive concepts or defined by means of more ultimate basic terms.

Is that an unwelcome conclusion? It was certainly not distressing to Immanuel Kant. In the (1781/1787) *Critique of Pure Reason*, Transcendental Analytic (First Division of the Transcendental Logic), under the Analytic of Concepts, Clue to the Discovery of all Pure Concepts of the Understanding, in the Table of Categories (A79-81/B105-107), Kant famously classifies causality as a concept of the pure understanding *Of Relations* incapable of analysis as to its origins in human thought. We began the discussion of causal relevance and relevant causation with a nod to David Hume, which makes it appropriate now near the end to turn to Kant, who reports in his (1783) *Prolegomena to any Future Metaphysics that can Come Forth as Science*, that it was Hume's scepticism about the necessity of causal connections that awoke him from his Wolffian–Leibnizian 'dogmatic slumber'. Kant [20] describes causality as among the 'original pure concepts of synthesis that the understanding contains within itself *a priori*' (A80/B106).⁹ If causation is a concept of the understanding, then it is part of the mind's innate framework by means of which it grasps phenomena generally in the

⁹Kant also writes:

Nature, considered merely as nature in general, is dependent upon these categories as the original ground of its necessary conformity to law (*natura formaliter spectata*). Pure understanding is not, however, in a position, through mere categories, to prescribe to appearances any *a priori* laws other than those which are involved in nature in general, that is, in the conformity to law of all appearances in space and time. Special laws, as concerning those appearances which are empirically determined, cannot in their specific character be derived from the categories,

world. It follows then, if Kant is right, that causation, like time and space as pure forms of intuition in the Transcendental Aesthetic of the first *Critique*, cannot be defined or reductively characterized in terms of more primitive ideas, let alone derived empirically or pragmatically from the contents of experience.¹⁰

Whether or not we are card-carrying Kantians, the circularity arguments demonstrating the irreducibility of relevant causation and hence of causation *simpliciter* are useful in pinpointing exactly how and where the effort to analyse or reductively define the concept of causation fails. Setting aside Kant's apparatus of transcendental idealism and architectonic categories of the pure understanding, we recognize that causation is a primitive concept as a result of the circularities that appear to be involved in explicating relevant causal connections on pragmatic grounds. Responding in this Kantian way to the *post hoc, propter hoc* fallacy leads philosophy an important step forward in mapping the hierarchy of metaphysical concepts and their logical and explanatory interrelations. If regarding causation as conceptually primitive, indefinable or unanalysable is somehow unpalatable, then the only alternative is to find a more plausible way to crack the circularity by which causation logically depends on causal relevance as a species of relevance, and where relevance generally and causal relevance in particular logically depend on the pragmatics of relevantly caused closure of cognitive agendas.¹¹

References

- [1] G. E. M. Anscombe. Causality and determination. In *Causation and Conditionals*, E. Sosa, ed., pp. 63–81. Oxford University Press, 1975.
- [2] T. Beauchamp and A. Rosenberg. *Hume and the Problem of Causation*. Oxford University Press, 1981.
- [3] J. Bennett. Event causation: the counterfactual analysis. *Philosophical Perspectives*, **1**, 367–386, 1987.
- [4] J. Carroll. *Laws of Nature*. Cambridge University Press, 1994.
- [5] J. D. Collins, N. Hall and L. A. Paul, eds. *Causation and Counterfactuals*. The MIT Press, 2004.
- [6] D. C. Dennett, ed. Cognitive wheels: the frame problem in artificial intelligence. In *Brainchildren: Essays on Designing Minds*, pp. 181–205. The MIT (Bradford Books) Press, 1998.
- [7] P. Dowe. *Physical Causation*. Cambridge University Press, 2000.
- [8] P. Dowe. A counterfactual theory of prevention and “Causation” by omission. *Australasian Journal of Philosophy*, **79**, 216–226, 2001.
- [9] P. Dowe and P. Noordhof, eds. *Cause and Chance: Causation in an Indeterministic World*. Routledge, 2004.
- [10] E. Eells. *Probabilistic Causality*. Cambridge University Press, 1991.

although they are one and all subject to them. To obtain any knowledge whatsoever of these special laws, we must resort to experience; but it is the *a priori* laws that alone can instruct us in regard to experience in general, and as to what it is that can be known as an object of experience (B165).

¹⁰Conceptual primitivism in the theory of causation is explored on different grounds among others by Anscombe [1] and Tooley [30, 31]. See also Carroll [4] and Menzies [23].

¹¹I have been positively and negatively inspired throughout this essay by Hitchcock [15]. Earlier versions of this essay were presented at the Conference on the Metaphysics of Science, University of Melbourne, Australia, July 2–5, 2009, and at the Second Conference of the European Philosophy of Science Association, Amsterdam, NL, October 21–24, 2009, under the title ‘Causal Relevance in the Metaphysics of Scientific Explanation.

- [11] J. H. Fetzer. The frame problem: artificial intelligence meets David Hume. In *Reasoning Agents in a Dynamic World: The Frame Problem*. K. M. Ford and P. J. Hayes, eds, pp. 55–69. JAI Press, 1991.
- [12] K. M. Ford and P. J. Hayes, eds. *Reasoning Agents in a Dynamic World: The Frame Problem*. JAI Press, 1991.
- [13] D. M. Gabbay and J. Woods. *Agenda Relevance: A Study in Formal Pragmatics*, Vol. 1 of *A Practical Logic of Cognitive Systems*. North-Holland (Elsevier), 2003.
- [14] T. Handfield, C. R. Twardy, K. B. Korb, and G. Oppy. The metaphysics of causal models. *Erkenntnis*, **68**, 149–168, 2008.
- [15] C. Hitchcock. A generalized probabilistic theory of causal relevance. *Synthese*, **97**, 335–364, 1993.
- [16] L. R. Horn and G. Ward, eds. *The Handbook of Pragmatics*. Blackwell Publishing, 2005.
- [17] D. Hume. *A Treatise of Human Nature* (1739-40), 2nd edn. In L. A. Selby-Bigge. Oxford University Press, 1985.
- [18] D. Jacquette. Review of Gabbay and Woods’s *Agenda Relevance*. *Studia Logica: An International Journal for Symbolic Logic*, **77**, 133–139, 2004.
- [19] A. Janiak. Newton and the reality of force. *Journal of the History of Philosophy*, **45**, 127–147, 2007.
- [20] I. Kant. *Critique of Pure Reason*. N. K. Smith, trans. St. Martin’s Press, 1965.
- [21] D. Lewis. Causation. *The Journal of Philosophy*, **70**, 556–567, 1973.
- [22] J. L. Mackie. *The Cement of the Universe: A Study of Causation*. Oxford University Press, 1980.
- [23] P. Menzies. Probabilistic causation and the pre-emption problem. *Mind*, **105**, 85–117, 1996.
- [24] M. Ramachandran. A counterfactual analysis of causation. *Mind*, **106**, 263–277, 1997.
- [25] A. Rosenberg. Propter hoc, ergo post hoc. *American Philosophical Quarterly*, **12**, 245–254, 1975.
- [26] E. Sosa, ed. *Causation and Conditionals*. Oxford University Press, 1975.
- [27] D. Sperber and D. Wilson. Fodor’s frame problem and relevance theory. *Behavioral and Brain Sciences*, **19**, 530–532, 1996.
- [28] D. Sperber and D. Wilson. *Relevance: Communication and Cognition*. Wiley-Blackwell, 1996.
- [29] D. Sperber and D. Wilson. Relevance theory. In *The Handbook of Pragmatics*, L. R. Horn and G. Ward, eds, pp. 607–632. Blackwell Publishing, 2005.
- [30] M. Tooley. *Causation: A Realist Approach*. Clarendon Press, 1987.
- [31] M. Tooley. Probability and causation. In *Cause and Chance: Causation in an Indeterministic World*, P. Dowe and P. Noordhof, eds, pp. 77–119. Routledge, 2004.
- [32] W. S. Wilkerson. Simulation, theory, and the frame problem. *Philosophical Psychology*, **14**, 141–153, 2001.

Received 29 August 2008