Causation

Causation

Helen Beebee

Please do not cite this version. The published version is:
'Causation', in B. Dainton & H. Robinson (eds), *The Bloomsbury Companion to Analytic Philosophy* (London: Continuum, 2013)

1. Introduction

The concept of causation pervades our ordinary talk and thought about the world. We routinely want to know what caused a given phenomenon, whether it is a car crash, a broken washing machine or a friend's bad mood, or what the likely effects of a given phenomenon will be. (What effect on house prices would a hike in interest rates have? Would a trip to the cinema improve my friend's mood?) And – as Elizabeth Anscombe (1971) famously pointed out – very many transitive verbs enshrine the concept of causation: to break something is to cause it to be in a broken state, to hurt someone is to cause them pain, and so on.

The concept of causation has also – at least in the last half-century or so – pervaded philosophical theorising. There are, for example, causal theories of knowledge (roughly: to know that *p* is to have been caused to believe that *p* by the obtaining of the fact that *p*; see e.g. Goldman 1967), perception (roughly: to perceive *o* is to have been caused to have an experience of *o* by the presence of *o*; see e.g. Grice 1961), rational decision-making (roughly: rational decision requires one to choose one's action on the basis of what the desirability of the likely effects of that action; see e.g. Lewis 1981), moral value (the goodness or badness of an act is determined by its consequences), and so on.

The concept of causation therefore seems to be an extremely important one (though, as we shall see, this claim has been disputed). But what is it that we attribute to a pair of events – my dropping a cup, say, and its breaking – when we claim that the first is a cause of the second? For it to be true that *a* caused *b*, it needs to be true not merely that *a* and *b* both happened, or that *a* happened (just) before *b*. Just after I dropped the cup, the telephone rang – but *that* was not caused by my dropping the cup. So what more is needed? As a first pass, we might try to say that *a* and *b* need to be *connected* in some way: perhaps we might say that *a produced b,* or *made b*

happen, or *brought b about*. But it seems that we are really just trading synonyms here; none of these ways of characterising the causal connection sheds any real light on its nature in the absence of further analysis.

This central question concerning its nature has dominated the extensive literature on the topic of causation in analytic philosophy in the last fifty years or so, and §§3 and 4 of this chapter explore some of the answers that have been proposed. In §2, I set the context for these answers with a very brief and partial account of two earlier contributions to the debate by David Hume and Bertrand Russell. In §3, I briefly rehearse one of the central points of dispute, namely whether causation should be seen in a 'Humean' or 'anti-Humean' light. In §4, I survey some of the most frequently discussed kinds of theory of causation. Finally, in §5 I discuss a selection of related issues concerning causation. What exactly does the relation of causation relate? (Events? Facts? Objects? All of the above?) Can absences or omissions be causes? Is the concept of causation univocal, or is there really more than one concept of causation? I also briefly discuss the 'causal exclusion problem'.

2. Hume and Russell

As an empiricist, Hume holds that every 'idea' (roughly: concept) has its source in an 'impression' – an experience of some kind. Having understood the importance of the idea of causation – it is, he says, the only relation that allows us to 'go beyond what is immediately present to the senses, either to discover the real existence or the relations of objects' (1739-40, 73) – he spends a large part of Book I of his *Treatise on Human Nature* searching for its elusive impression-source. He identifies three core components of the idea of causation, namely priority (causes precede their effects), contiguity (causes and effects are contiguous, or right next to each other, in space and time) and 'constant conjunction' (causes and effects are constantly conjoined, i.e. if *a* causes *b*, then whenever an event similar to *a* occurs, it is followed by an event similar to *b*). But, he insists, these three conditions are not enough: our idea of causation includes the idea of 'necessary connection'. After all, if the relation of causation is to inform us 'of existences and objects, which we do not see and feel' – that is to say, if it is to underpin our *inferences* from causes to effects (as when I drop the cup and infer that it will shortly break) – then it would seem that causes must *guarantee* the occurrence of their effects, or, in other words, they must *necessitate*

their effects. And so the impression-source of the idea of necessary connection must be found if the idea of causation is to be legitimised.

As it turns out, Hume eventually turns this last thought on its head: rather than our inference from cause to effect being underpinned by our responding to an objective necessary connection between the two, it turns out that the inference itself is the impression-source of the idea of necessary connection. Once I have experience of *A*s being constantly conjoined with *B*s, on observing an *A* I infer that a *B* will follow – not through some sophisticated piece of reasoning, but through 'Custom or Habit' (1748/1751, 43), or by what psychologists nowadays call 'associative learning'. So there is a 'transition' in the mind from the experience of an *A* to the expectation that a *B* will follow. This transition itself produces an impression, and that impression, Hume claims, is the source of the idea of necessary connection. Contrary to what we might have thought, then, the necessity that is part and parcel of our concept of causation is supplied not by the world, but by our own minds; indeed, the supposed notion of a 'nexus' between cause and effect that is independent of the mind is unintelligible, since it lacks an impression-source.

For present purposes, Hume's account of causation raises two central and closely connected sets of questions. First, does causation somehow or other involve *necessitation* – do causes *make* their effects happen? If so, what is the nature of this necessity? Is it something supplied by the mind, or is it instead a feature of mind-independent reality – and, if the latter, is such a thing genuinely intelligible? Second, does causation require 'constant conjunction' or *regularity*? Is it essential to our concept of causation that the same cause will always and everywhere be followed by the same effect? If so, can this fact be used to shed light on the nature of causation?

These two sets of questions are taken up by Bertrand Russell in his 'On the notion of cause' (1912-13). Russell begins by complaining that 'the word "cause" is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable' (1912-13, 1), and, famously, noting that the 'law of causality' – roughly, the principle of 'same cause, same effect' – 'is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm' (*ibid.*). Russell broadly agrees with Hume on the unintelligibility of the notion of some mind-independent nexus between causes and effects. But he is equally dismissive of Hume's positive contention that causation requires contiguity, temporal priority and constant conjunction. In particular, he

argues that the regularities we observe in daily life are almost all merely 'fairly dependable' rather than exceptionless (1912-13, 8); and once we appeal to the sciences to supply our supposed exceptionless regularities, what we find is that once we have specified the full cause of a given event in sufficient detail to render our regularity exceptionless, we'll find ourselves with a circumstance that is unlikely ever to be repeated and hence cannot be considered to be an instance of a 'regularity' at all.

Perhaps the most enduring aspect of Russell's critique of the notion of cause, however, is his claim that 'in advanced sciences such as gravitational astronomy, the word "cause" never occurs' (1912-13, 1). For Hume, the relation of causation is central to our ability to draw inferences beyond what we can 'see or feel'; for Russell, our best science, namely physics, has 'ceased to look for causes' because 'there are no such things' (*ibid,*); and physics is none the worse – indeed it is all the better – for that.

3. Humeanism and its critics

Let's return to the two sets of questions posed above, concerning the idea that causes necessitate their effects and the idea that causation and regularity are intimately related. One way to answer these questions – and this is a view that has been attributed to Hume by many commentators – is to claim that causation just *is* a matter of regularity: for *a* to cause *b* just *is* for all events like *a* (the *A*s) to be followed by events like *b* (the *B*s). So *a* causes *b* if and only if all *A*s are followed by *B*s. This view is sometimes called the 'naïve regularity theory' of causation. According to the naïve regularity theory, there is no real necessity involved in causation at all. Or, to put it another way, while it is true that 'all *A*s are followed by *B*s' and 'an *A* occurs' *entail* that a *B* will follow, and entailment is a species of necessity, the 'necessary connection' between *a* and *b* obtains merely in virtue of the fact that all *A*s are followed by *B*s: there is no intrinsic relation of necessity, or indeed a nexus of any kind, between *a* and *b*.

Unfortunately the naïve regularity theory is subject to an insurmountable battery of objections, of which I shall mention just two: the problem of accidental regularities and the problem of the common cause. There are some truths of the form 'all *A*s are followed *B*s' where it is manifestly not the case that *A*s *cause B*s. Starting with the problem of accidental regularities: imagine that on all four occasions when I took a bath in my previous house (I very rarely take a bath) (*A*), the neighbour rang

the door bell (*B*) just as I stepped in. So all *A*s have been followed by *B*s so far, and since I have now moved house, all the *A*s there will ever be are followed by *B*s. Of course, it's *possible* that the neighbour somehow knew I was taking a bath and deliberately rang the bell to annoy me – in which case, my taking a bath *would* have been a cause of her ringing the bell. But according to the naïve regularity theory, my taking a bath was automatically a cause of the neighbour's ringing the bell, just in virtue of the fact that the former was always followed by the latter. Clearly that's not right: it's just an accident that the regularity holds. The problem of the common cause is similar. Imagine that every time a particular barometer points to 'rain', it rains shortly afterwards. According to the naïve regularity theory, the position of the barometer pointer causes the rain. But manifestly this is not so: the position of the barometer pointer and the rain are effects of a common cause, namely low air pressure.

Despite the obvious failings of the naïve regularity theory, many philosophers have attempted to provide more sophisticated 'regularity theories' of causation, which seek to ground facts about causation in facts about regularities but conceive the relationship in more complex way that avoids the problems that beset the naïve version (see for example Mackie 1965). The version of the regularity theory that has received the most attention, and perhaps the one with the best chance of success, is the counterfactual analysis of causation first developed by David Lewis (1973) and discussed in §4 below.

Regularity theories of causation are often referred to as 'Humean' theories not just because they preserve the spirit of the naïve regularity theory that is often attributed to Hume, but also for the deeper reason that they eschew the kind of nexus or mind-independent necessity that Hume claimed to be unintelligible. The eschewal of necessity that has driven some philosophers in the direction of a regularity theory also no doubt gained some momentum from W. V. Quine, according to whom '*de re*' necessity (that is, all necessity that does not ultimately derive from the meanings of words, as in 'necessarily, all bachelors are unmarried') is deeply suspicious and therefore – echoing Russell – has no place in philosophical discourse (Quine 1963).

It is worth remembering, however, that Hume himself – unlike contemporary regularity theorists – apparently endorses the claim that the idea of necessity is an essential part of the idea of causation; it's just that the necessity involved in causation turns out contributed by the mind rather than the world (see Beebee 2006, Chs.4 & 7).

Other kinds of broadly Humean theory stay closer to what may have been Hume's real view in this regard. For example, Peter Menzies and Huw Price (1993) argue that our concept of causation has its roots in our experience not as predictors, as Hume had it, but as agents. Price (2007) later characterises causation as a 'perspectival' phenomenon. The world exhibits a fundamental asymmetry in that entropy tends to increase over time; and it is because of this asymmetry, Price thinks, that we think of the future as 'open' and the past as 'fixed'. Thus when we deliberate we hold fixed what has already happened and regard what has yet to happen as up to us, to be determined (in part) by what we decide to do. The deliberative perspective designates the actions we might perform (such as my putting on the kettle) as means and events that lie in the future (such as my getting my desired outcome of a cup of tea), but not events that lie in the past (such as my having had a cup of tea earlier in the day) as ends; and it is in this deliberative perspective that he locates our conception of the world as a world of causes (means) and effects (ends). A related view is held by Jon Williamson (2006), who develops an 'epistemic' theory of causation, according to which – again, in broadly Humean spirit – causation is conceived in terms of what beliefs it would be rational for us to adopt for the purposes of predicting, controlling and explaining what goes on in our environment: 'the causal relation is characterised by the causal beliefs that an omniscient rational agent should adopt' (2006, 7), where the rationality of those causal beliefs is secured by the right kinds of evidential relations rather than their latching onto some mind-independent causal nexus. (See also Ramsey 1929.)

It is, however, open to question whether we should agree with Hume about the alleged unintelligibility of necessary connections, and plenty of philosophers simply reject Hume's claim, thereby constituting the 'anti-Humean' wing of the debate. For example, Adrian Heathcote and David Armstrong (1991) propose an account of causation that derives from Armstrong's (1983) theory of laws of nature. On Armstrong's account of laws, a law of nature is a relation of necessity (N) between universals, so that it is a law that all Fs are (or are followed by) Gs – where F and G are universals – just in case $N(F, G)$. The basic idea is that what makes it a *law* that all Fs are followed by Gs, as opposed to its merely being *true* that all Fs are followed by Gs, is precisely that in the former case, but not the latter, F and G are related by N. (So – to use toy examples – there is no necessary connection between taking a bath and the neighbour ringing the doorbell; but there *is* a necessary connection between

being an object with a particular mass *m* subject to force *f* and accelerating at rate *a*: any object possessing the universals *m* and *f must* accelerate at rate *a,* where *f=ma*.) Armstrong and Heathcote hold that causation is simply the instantiation of a law, thus conceived. So while the law relation *N* relates universals *F* and *G*, say, that law is instantiated in particular instances or 'states of affairs', so that a *particular* object's having *F* and its having *G* – those two states of affairs – are themselves related by an instance of *N*.

More generally, Hume's claim about the unintelligibility of mind-independent necessity is grounded in a version of empiricism that is no longer widely held. As we saw in §2, according to Hume all 'ideas' must be derived from 'impressions'; in other words, to put it crudely, you can't have a concept of something that you have no experience of (unless you can somehow construct the concept out of materials you *do* have experience of – so for example we can form the idea of a unicorn by combining the idea of a horse with the idea of a horn, both of which we have experience of). But this version of empiricism is widely rejected as too demanding. In particular it seems to have unpalatable consequences for the kinds of unobservable entity that are commonplace in scientific theories: quarks, electrons, forces, fields, and so on. Since we can have no direct experience of such entities, Hume's empiricism seems to entail that we cannot so much as form concepts of them – which makes it rather hard to see how to make any sense of contemporary physics. These days, many philosophers count themselves as empiricists in a more relaxed sense that allows us to have legitimate concepts of unobservable entities, so long as those entities have a clearly definable role within the theoretical framework of the sciences. And we might extend that view to cover more recherché items of ontology such as Armstrong's *N* or, more generally, an intrinsic causal relation or a 'nexus' between causes and effects (see Menzies 1998).

A second, more direct challenge to Hume simply rejects his claim that necessity is not observable. Elizabeth Anscombe, for example, famously says: 'Hume 'confidently challenges us to "produce some instance, wherein the efficacy is plainly discoverable to the mind, and its operations obvious to our consciousness or sensation" [Hume 1739-40, 157-8]. Nothing is easier: is cutting, is drinking, is purring not "efficacy"?' (1971, 93). My own view here is that facts about our experience (e.g. the fact that have visual experience as of one billiard ball making another one move) simply do not settle the question about whether *what* we are

observing is the kind of 'nexus' between causes and effects to which Hume was so hostile (see Beebee 2003 and 2009).

Whatever the truth of the matter concerning the correct interpretation of Hume, and whether his argument for the unintelligibility of a mind-independent necessary connection between causes and effects is any good, it is certainly true that Hume has had an enormous influence on the shape of the contemporary debate about the nature, and our understanding, of causation.

4. Some theories of causation

This section provides an admittedly brief and partial survey of some recent theories of causation. Some theories have already been mentioned: the naïve regularity theory, the counterfactual theory, Price's perspectivalism, Williamson's epistemic view, and Heathcote and Armstrong's overtly anti-Humean position. Of these, only the counterfactual theory is discussed further below. Added to the mix in this section are 'process' and 'mechanistic' positions and probabilistic theories of causation.

Counterfactual theories of causation

The fundamental insight behind counterfactual theories of causation is the very simple thought that effects *depend* upon causes in some way; and counterfactual theories cash out the notion of dependence in terms of *counterfactual* dependence. Suppose I strike a match (*c*), thereby causing it to light (*e*). Then – with some background assumptions in place, of which more later – if, contrary to what in fact happened, I hadn't struck the match, it wouldn't have lit. In other words, *e counterfactually depends on c*. By contrast, I step in the bath (*d*) and the neighbour rings the doorbell (*f*). If I hadn't stepped into the bath, the neighbour would still have rung the doorbell: *f* does not counterfactually depend on *d*. This suggests a straightforward analysis of causation, as follows: *c* causes *e* if and only if had *c* not occurred, *e* would not have occurred either.

Unfortunately, things aren't so straightforward. First, we have now replaced one mystery with another: without an account of how facts about counterfactual dependence are determined, it's not clear that we have made very much progress. (I return to this issue below.) Second, the account just proposed cannot be right because it is subject to obvious counter-examples. In particular, it fails in cases of *pre-emption*, where there is a back-up cause waiting in the wings to cause the effect,

should the actual cause fail to occur. Imagine two assassins, A1 and A2, both trying to shoot and kill the President – one from the grassy knoll, and the other from the book depository window. A2, up in the book depository, has taken aim is about to shoot when out of the corner of her eye she sees A1, out there on the grassy knoll, take aim and fire. Knowing that A1 is a crack shot (and she is right: A1 shoots and kills the President), A2 lowers her gun and makes her escape. A1's shot (*c1*) caused the President's death (*e*). But if *c1* hadn't happened, A2 would have fired instead (*c2*), and *e* would still have happened. In other words, if *c1* hadn't happened, *e* would still have happened. But that's inconsistent with the simple analysis proposed above, since clearly *c1* was a cause of *e*. Problem.

Lewis's (1973) solution to the problem is to hold that causal *dependence* is a matter of counterfactual dependence, and that causation is a matter of a *chain* of causal dependence relations. Take some event that is on the causal path from A1's shot to the politician's death: A1's bullet piercing his heart, say. (Call this *d*.) If *c1* hadn't happened, *d* wouldn't have happened: if A1 hadn't fired her gun, her bullet never would have got there. So *d* causally depends on *c1*. And *e*, in turn, causally depends on *d*, since if the bullet hadn't pierced his heart (let's suppose), he wouldn't have been killed – our back-up assassin having given up by this point and begun to make her escape. So there is a chain of causal dependence from *c1* to *e* via the intermediate event *d*; hence *c1* is a cause of *e*, which is the right answer.

So far, so good. But there is a host of other problems lurking. One is the problem of 'late pre-emption'. Imagine instead that assassin A2 has been given strict instructions not to desist unless and until she sees the President well and truly dead. In that scenario, she's still on the scene with her gun trained on him at every point in between A1 firing and the President dying. So there is no event *f* in the causal chain between *c1* and *e* such that *e* counterfactually depends on *f*. For example, if A1's bullet hadn't pierced his heart (and so, we may suppose, had not fatally injured him), then the President would have died anyway because A2 would still have shot him. (Late pre-emption is so-called because the event that stops the back-up cause from kicking in is the effect itself – in this case, the President's death – whereas in 'early' cases some much earlier event does the job; for example in the first case above it is A1's taking aim that stops A2 from firing.)

Another problem is that of 'trumping pre-emption' (Schaffer 2000a). Imagine that there are wizards who can cast spells, and that those spells can work at a spatial

and temporal distance with no intermediate states or events that 'join up' the casting of the spell with its end result. Merlin casts a spell at midnight to turn the prince into a frog, and Morgana does likewise just afterwards. Imagine further that it is built into the laws of nature that if two spells are cast with the same intended effect, the later spell always cancels the first one, so that it is the second spell that is effective. So it is Morgana's spell (*c*), and not Merlin's, that causes the prince to turn into a frog (*e*); but – since this is a case of pre-emption – there is no counterfactual dependence of *e* on *c*: had *c* not occurred, *e* would have happened anyway, caused by Merlin's earlier spell. By stipulation there is no intermediate event *f,* such that *d* counterfactually depends on *c* and *e* in turn counterfactually depends on *d*. So the counterfactual analysis cannot account for the fact that *c* caused *e*. Of course, there are in fact no wizards and spells; but an analysis of causation is supposed to tell us how the causal facts are determined in every *possible* situation, not merely in every *actual* situation.

Many attempts have been made to come up with a counterfactual analysis of causation that isn't susceptible to these problems; see for example Noordhof 1999. Lewis's own attempt (2000) involves appealing to the notion of 'influence'. Roughly, the idea is that rather than conceiving causation as a matter of *whether* a given effect would have occurred had the cause not occurred, but rather a matter of the extent to which the effect would have been different had various 'alterations' of the cause occurred. For example, in our 'late' assassin case, while it's true that the President would still have died had A1 not fired, the time of his death is still sensitive to the precise time at which A1 fired: had A1 fired a little earlier or later, the President's death would correspondingly have occurred a little earlier or later. So A1's firing has enough 'influence' over the President's death for the former to count as a cause of the latter.

Counterfactual theories and Humeanism

In §3, I described counterfactual theories of causation as 'Humean'. Why is that? Well, notice that I have not yet said anything about what determines the truth of the counterfactuals ('had I not struck the match, it wouldn't have lit', and so on) that in turn are supposed to determine the truth of causal claims. And it is in Lewis's analysis of counterfactuals that we discover the Humean credentials of counterfactual theories of causation.

Lewis provides a *possible world analysis* of counterfactuals. Think of the whole of the Universe as the 'actual world'. It is the way it is – but there are many ways it *might* have been. For example, the laws of nature might have been different. Or the laws of nature might have been the same but the Universe might have started out with different initial conditions, and so – even with the same laws – things might have panned out very differently. Think of all of the ways the world *might* have been as ways some merely *possible* world really *is*.

Now, some possible worlds are more similar, or 'closer' (in a metaphorical sense), to the actual world than others. A possible world that is exactly similar to the actual world for the first 6 billion years of its history but then starts to diverge in similarity is much closer to the actual world than is one that started out radically different to the actual world and remains so throughout its history. More generally, Lewis (1979) holds that the closeness of one possible world to another is determined by two things: sameness of 'matters of particular fact' (roughly: what happens) across regions of spacetime (a bigger region of perfect match makes for greater similarity), and sameness of the laws of nature. And Lewis's theory of laws of nature is itself a sophisticated 'regularity theory' of laws, namely a 'best system' account: 'a contingent generalization is a *law of nature* if and only if it appears as a theorem (or axiom) in each of the true deductive systems that achieves a best combination of simplicity and strength' (1973a, 73). To put it more crudely, for Lewis the laws of nature will turn out to be subset of the regularities: they will be those regularities that are the most powerful for making predictions.

With this account of closeness in place, Lewis's account of counterfactuals runs as follows. Let '*~O(c) □→ ~O(e)*' mean: 'if event *c* had not occurred, event *e* would not have occurred'. To evaluate the truth of this counterfactual (we are assuming that both *c* and *e* did occur in the actual world), we consider the *closest* possible world *w* at which *c* does not occur – that is, the closest world at which '*~O(c)*' is true – and see whether or not '*~O(e)*' is true at *w*. If it is, then the counterfactual is true; if it isn't, the counterfactual is false. So for example suppose our counterfactual is 'if I hadn't struck the match at *t*, it wouldn't have lit'. What is the closest possible world at which I don't strike the match (*w*) like? Lewis's answer is: it is a world that perfectly matches the actual world throughout the whole of spacetime until just prior to *t* (until *t*-1, let's say), and which has exactly the same

laws as the actual world *except* that at *t*-1 a 'small miracle' occurs: there is a minor violation of the actual world's laws – just enough of a violation to accommodate my failing to strike the match (since, we may suppose, in the actual world the laws determine that I *do* strike the match). Thereafter, *w* continues to evolve according to the same laws as those that hold in the actual world. (What happens after *t*-1 at *w* will of course start to diverge – and increasingly so – from what happens at the actual world. In particular, the match – unstruck at *w* – will not light, since, given the sameness of *w*'s laws and the actual laws (with the exception of the miracle required to stop me striking the match), unstruck matches do not spontaneously light of their own accord at *w* any more than they do at the actual world.)

With all this in place, we are in a position to see why Lewis's counterfactual analysis of causation counts as a 'Humean' theory. Counterfactuals are to be analysed in terms of what happens at the closest possible worlds, and closeness of worlds is in turn determined by (a) similarity in what happens and (b) sameness of the laws. But the laws themselves are merely a subset of the regularities. Thus, ultimately, the truth of '*c* caused *e*' is determined by the overall pattern of what happens at some close possible world, and not by the existence of any causal 'nexus' between *c* and *e* in the actual world. Of course, Lewis's theory of laws is not obligatory: for example one might, as we have already seen, hold that laws of nature are a matter of the obtaining of a necessitation relation between universals. However it is not at all clear that such an 'anti-Humean' account of laws is really compatible with Lewis's analysis of counterfactuals. And even if it is, it remains true that Lewis's *own* conception of the counterfactual analysis of causation is intended to be broadly Humean one, given that closeness of worlds is, for Lewis, determined solely by 'Humean' facts.

Processes and mechanisms

Worries that Lewis's revised counterfactual analysis (2000) simply falls foul of a new set of counterexamples have been well voiced in the literature (see e.g. Schaffer 2001). A more general question to ask, however, is whether the counterfactual analysis really gets to the heart of the nature of the causal relation. One can think of Lewis's revised account as a kind of 'black box' account: we 'wiggle' the input (altering the time or manner of assassin A1's shot, say) and see what, if anything, happens to the output (the time or manner of the President's death). But isn't what's really important, as far as causation is concerned, what's going on in between the

two? After all, it's just *obvious* that A1's shot causes the death, because we can easily trace the causal path from one to the other via the expulsion of the bullet from the gun, its movement towards the victim, its entering the victim's heart, and so on. If the counterfactual analysis has such difficulty accounting for the causal relation between the shot and the death, perhaps that shows us that we are looking in the wrong place for an analysis of causation. Perhaps we should instead be looking inside the black box, as it were, to trace out the very obvious *process* or *mechanism* that connects the shot with the death.

This kind of approach has been taken by a number of philosophers. Perhaps the most well worked-out 'process' view is that of Phil Dowe (2000), building on earlier work by Hans Reichenbach (1928), Jerrold Aronson (1971), David Fair (1979) and Wesley Salmon (1984). Reichenbach is concerned with the difference between real and 'pseudo-' processes. For example, as I take a walk on a sunny day, my movement along the street constitutes a genuine causal process, with later stages causally depending on earlier stages. By contrast, the movement of my shadow is a mere pseudo-process. The later position of my shadow is not caused by its earlier position; rather, each is independently caused by my own position. How do we account for this difference? Reichenbach's answer (1958) is that only real processes can 'transmit a mark': intuitively, if one were to modify my walk by means of a single intervention – say by me picking up a large box en route – that 'mark' would be transmitted to later stages of the process (so long as I don't put the box down again). By contrast, if one directly modified the *shadow* at any point (e.g. as I walk past a bus stop, the shadow acquires a bus-stop-shaped modification), that mark would *not* be transmitted to later stages: the bus-stop-shaped modification goes away again as soon as I've passed by.

Salmon (1984) refines Reichenbach's criterion for distinguishing real from pseudo-processes; however, his view has also been subjected to apparently decisive objections (see Dowe 2008, §6 for a summary). Dowe's 'Conserved Quantity Theory' of causation (2000) is a version of the process theory that aims to overcome these worries. Roughly, the idea is that what is distinctive of causal processes is not that they transmit a *mark*, but that they transmit a *conserved quantity*, where a conserved quantity is any quantity identified by the laws of nature as universally conserved – according to our current best science, these would be charge, linear momentum and mass-energy. So for example the movement of a billiard ball across the table involves

the transmission of mass-energy and linear momentum from earlier to later stages, whereas a shadow is not the kind of object that can so much as possess a conserved quantity, and hence no conserved quantity can be transmitted from earlier to later stages of the shadow's movement.

One worry about the Conserved Quantity Theory (see Dowe 2000, Ch.1 and Lewis 2004) is that at best it only tells us what causation *actually* consists in; it does not tell us anything about *why* the transmission of conserved quantities is what is picked out by our concept of causation. Analogy: suppose that all mental states – pain, for example – are in fact brain states (the firing of C-fibres, say). If you were an alien who knew nothing about pain, and were told that the concept *pain* picks out the firing of C-fibres, you would still be none the wiser about why *these* particular brain states are picked out by the concept of pain: you'd know which brain states constitute pain states, but you wouldn't know what it *is* to be in pain. In order to know *that*, you'd need to have something like a conceptual analysis of pain: you'd need to realise that we give the name 'pain' to whatever brain states are (say) typically caused by bodily damage of some kind (being pricked with a pin or stubbing one's toe or putting one's hand in a flame, etc.) and which in turn typically cause certain kinds of behaviour (removing oneself from the source of damage as quickly as possible, saying 'ouch!', etc.). And the firing of C-fibres (we're imagining) is what is picked out by the concept *pain* precisely because it is the state that has just those typical causes and effects: it is the physical state that 'plays the pain role', as it is sometimes put. (This kind of account of mental states is known as 'functionalism'; see Levin 2010.)

The Conserved Quantity Theory tells us what causation is, in the same sense that neuroscience might tell us that that pain is the firing of C-fibres. It tells us nothing, however, about why the transmission of conserved quantities 'plays the causation role': it doesn't tell us what it *is* for one thing to cause another. Peter Menzies (1996) has attempted to plug this gap by specifying the role that the transmission of conserved quantities plays: he proposes a 'folk theory' of causation (including 'platitudes' such as that causation is a relation between events, and that causes typically raise the probability of their effects) and suggests that the transmission of conserved quantities is that feature of the world (analogous to the firing of C-fibres in the case of pain) that in fact satisfies those platitudes.

More recently, some philosophers have offered 'mechanistic' rather than process-based accounts of causation (Glennan 1996; Machamer, Darden and Craver

2000). Machamer *et al* note that 'terms like "cause" and "interact" are abstract terms that need to be specified with a type of activity and are often so specified in typical scientific discourse. Anscombe [1971] … noted that the word "cause" itself is highly general and only becomes meaningful when filled out by other, more specific, causal verbs, e.g., scrape, push, dry, eat, burn, knock over. An entity acts as a cause when it engages in a productive activity' (Machamer *et al* 2000, 6). So the general idea is that the concept *cause* is merely an abstraction from the many different kinds of specific mechanism that we find in the world, and that are investigated by the sciences. So to uncover the nature of such a mechanism just *is* to discover the nature of one specific kind of causal relation. The mechanistic view can thus be seen as an attempt to explain Russell's observation that the word 'cause' is not to be found in the advanced sciences: such sciences have no need for such an abstract term when they can provide the nuts-and-bolts detail concerning the specific mechanisms they are concerned with.

One way to motivate a mechanistic account, as opposed to the Conserved Quantity Theory, is to note that the latter seems apt to capture causal processes visible from the perspective of physics, but is less obviously appropriate for the kinds of process studied by other sciences, such as molecular biology or the medical sciences. We need not dispute whether the transmission of conserved quantities is involved in interactions in, say, molecular biology; the point is merely that the mechanisms by which causal influence is transmitted are identified independently of consideration of conserved quantities. To use a more mundane example, you don't need to know anything about fundamental physics to be able to identify the mechanism by which my letter posted in Manchester arrives at its intended destination in London. (See Williamson 2011 for a survey of mechanistic accounts.)

Probabilistic theories

An approach to causation that appears, at least at first sight, to be very different to those described above takes its starting point from the thought that our causal talk and thought encompasses not merely causal claims about *particular* events or happenings, but also *general* causal claims: we say that smoking causes cancer, that eating sugary foods causes tooth decay, that lack of economic growth causes unemployment, and so on. From an epistemological point of view, such claims are rarely arrived at or justified by generalising over specific cases of 'particular' causation. For example, the process by which smoking causes lung cancer in any particular case is still not

especially well understood, and similarly it is impossible to answer the question whether or not any particular smoker who gets lung cancer *would* have got it had they not smoked with any great degree of confidence. And yet is virtually universally agreed that smoking causes lung cancer *in general*.

How is this possible? The answer, of course, is that such general causal claims are arrived at on the basis of statistical methods: very roughly, the incidence of lung cancer amongst smokers is very much higher than it is amongst non-smokers, and it is this general increase in the probability of lung cancer amongst the smoking population that justifies our belief that smoking causes lung cancer. In other words, where S = smoking and C = lung cancer, $Pr(C/S) > Pr(C/\sim S)$; and this is at least *prima facie* evidence that S is a cause of C.

Probabilistic theories of causation seek to analyse (or, more weakly, to shed some light on) the notion of 'general' or 'type-level' or 'population-level' causation by appealing to this kind of statistical fact (e.g. Suppes 1970, Eells 1990; see Hitchcock 2011 for an overview). A first pass at a theory might be: A causes B (in some specified population: adult males, or UK residents, or …) if and only if $Pr(B/A) > Pr(B/\sim A)$ – that is, if and only if *A raises the probability of B* in that population. Unfortunately, however, this crude analysis fails in both directions. First, there can be *spurious* statistical correlations. For example where E = a high level of educational attainment and P = attending a private school, $Pr(E/P) > Pr(E/\sim P)$. However, it turns out (this may be an urban myth, but it certainly *could* be true, so let's suppose it is) that the reason for this correlation is that children's educational attainment tends to match that of their parents and parents with a high level of attainment (A) are more likely to send their children to private schools. In other words, A is a *common cause* of E and P, and there is in fact no direct causal relation between E and P: going to a private school does not *cause* high educational attainment.

In common cause cases, if we 'hold fixed' the common cause – in this case A – and consider the statistical correlation between our two effects (E and P) separately in the presence and absence of A, the correlation between E and P disappears: $Pr(E/P \,\&\, A) = Pr(E/\sim P \,\&\, A)$, and $Pr(E/P \,\&\, \sim A) = Pr(E/\sim P \,\&\, \sim A)$. So this new correlation lines up with the fact that P is not a cause of A: in our new 'reference classes' or 'background contexts' (considering children with parents who have a high level of attainment and those without separately), P does not increase the probability of E. We cannot, however, simply amend our analysis of causation by stipulating that causes

raise the probability of their effects in reference classes where all other *causes* of the effect are held fixed, since this would be blatantly circular. One suggested remedy (Eells 1991) is simply to hold *all* other prior factors fixed, irrespective of whether or not they are causes of the effect under consideration. One consequence of such an analysis, however, would be to divorce the analysis of 'general' causation from its epistemology: in practice we never come across wholly homogeneous reference classes and so are not in a position to come to know what the statistical correlations in those reference classes are (see Dupré 1984, 1990; Eells 1987).

A related approach to general causation (see e.g. Pearl 2000; Woodward 2003), which sometimes goes under the heading of 'causal modelling', or 'manipulability' or 'interventionist' theories, effectively gives up on the project of providing a full-blown non-circular *analysis* of causation but nonetheless seeks to provide an illuminating account of how we can infer the causal structure of a given kind of situation from statistical evidence. The central question here is: what kinds of probabilistic relationship between a given range of variables (ranging over, say, different kinds of school, different levels of parental attainment, average income in the locality of the school, etc. if we are interested in the causes of levels of educational attainment) need to be in place in order for us to be able to infer what the *causal* relationship is between those variables?

Probabilistic theories generally can be thought of as broadly Humean theories, because they seek to analyse general causation in terms of statistical regularities. However interventionist theories in particular fail to provide a full-blown analysis of causation because they rely on the notion of a 'manipulation' or an 'intervention' – and this is itself a causal notion. Here, very roughly, is why. Take our educational attainment example again. Suppose we just looked at the statistical relationships between three variables: average income in the locality of the school, type of school (state or private, let's say) and average level of qualification of the teachers. If we did this, we might well find a statistical correlation between type of school and educational attainment. But – as we've already seen – this might not be a *causal* correlation: both might be effects of parental attainment, a variable that we have not included. In practice, of course, we cannot include *all* the variables that are relevant to attainment – it's just far too complicated. So how do we ensure, with only some of the relevant variables represented in our model, that the statistical and causal correlations line up? The answer is that we consider only the statistical correlations that arise from

interventions, where we 'fix' or 'set' the value of a given variable in a way that breaks any connection with prior causes that are outside the system. In our toy example, imagine that we don't simply consider the (spurious, in fact) correlation between going to private or state school on the one hand and higher or lower educational attainment on the other. Instead, what we do is take a bunch of randomly selected 11-year-olds and send them off to private school, and send a second randomly selected bunch of 11-year-olds off to state school, and compare their attainment levels 5 years later. Random selection will (unless we are very unlucky) break the correlation between parental attainment and kind of school attended, since we didn't pay any attention to parental attainment when deciding which children to send to which kind of school. And so we would expect there to be no correlation between kind of school attended and the level of attainment. In effect, what we have done here is 'intervened' on the variable *kind of school attended* in such a way as to screen off the effects of the usual causes of the value of that variable; and if we do *that*, the statistical correlations will line up with the causal correlations. Of course, in practice we cannot simply randomly select children and pack them off to one kind of school or another. However, this kind of practice is exactly what happens in standard randomised trials. When patients are randomly assigned to either trying out a new drug or else staying with the old one, say (without being told which group they belong to), the aim is precisely to control for other potential causes: if patients (or their doctors) make the choice themselves, spurious correlations may well emerge, just as they do if parents make the choice about which school to send their children to.

The notion of an 'intervention' is, however, a causal notion, since an intervention on X is *defined* in terms of its breaking the causal connection between X and its causes. Thus this kind of probabilistic theory fails to deliver a full-blown analysis of causation. Whether or not this is a problem depends, of course, on what one's aims are in providing a theory of causation in the first place; if one's aim is to find a full-blown analysis, than an account such as Woodward's interventionist account will not, just by itself, fit the bill. On the other hand, if one's aim is to gain a practically workable theory of how to extract causal information from complex statistical relationships, it fares considerably better than other theories of causation. (See Woodward 2008 for a general discussion of interventionist/manipulability accounts of causation.)

5. Further issues

§4 was a brief and incomplete survey of some of the most widely discussed theories of causation of recent years. In this final section, I discuss some more general issues surrounding causation that bear on the question of which theory (or theories) constitute our best account of the nature of causation.

What does causation relate?

I have been talking so far of causation (or at least 'particular' or 'token' causation, as opposed to general causation) as though it relates *events*. But what is an event – and are events the only candidates for the relata of the causal relation? Various accounts of the nature of events exist (see e.g. Davidson 1967, Kim 1973, Lewis 1986a); I shall here briefly discuss just Lewis's account. For Lewis (1986a), an event is a region of spacetime that has certain of its features essentially and some only contingently. Or rather, the same region of spacetime will have many Lewisian events occurring in it, all with different essential and accidental features. For example, in the region I am currently occupying, several events are currently occurring, one of which is essentially my drinking coffee, another is essentially my drinking coffee slowly, and another is essentially my drinking coffee out of a mug.

The more essential features an event has, the more 'fragile' it is (so my (essentially) drinking coffee slowly (*e*) is more fragile than my drinking coffee (*f*)) – in other words, the fewer possible worlds it occurs in. A possible world that is more or less like the actual world but where I am drinking my coffee quickly is a world where *f* occurs but *e* does not. The distinction between more and less fragile events is needed in order to make the counterfactual analysis of causation get the causal facts right. The reason I'm drinking my coffee slowly is that it is very hot – so its being hot is a cause of my drinking it slowly, but it is not a cause of my drinking it *simpliciter*. Likewise, my drinking the coffee slowly counterfactually depends on its being hot (if it had not been so hot, I would have been drinking it quickly), while my drinking the coffee *simpliciter* does not (if it had not been so hot, I still would have been drinking it).

Lewis imposes some restrictions on what can count as an event. In particular, events must not have 'overly extrinsic' essences. Again, this is to ensure that the counterfactuals line up with the causal facts: my buying coffee just now (*c*) is a cause of my now drinking it, but it is not, intuitively, a cause of my-drinking-my-coffee-or-

the-sun-exploding. If the latter were allowed to count as an event, then the counterfactual analysis would get the wrong answer, since it is true that had I not bought coffee just now, I would not now be drinking it, and nor would the sun have exploded. But my-drinking-my-coffee-or-the-sun-exploding would be an event whose essence is overly extrinsic; hence it is not really an event at all according to Lewis, and hence, since causation is a matter of counterfactual dependence between *events*, the counterfactual dependence of my-drinking-my-coffee-or-the-sun-exploding on *c* doesn't constitute a counterexample to the counterfactual analysis.

Others (see e.g. Bennett 1988, Mellor 1995) argue that the most basic form of causal claim is '*E* because *C*', where '*C*' and '*E*' state facts and 'because' is a sentential connective. So there can be causal claims that are true but not in virtue of the obtaining of a causal *relation*. One major reason why one might hold this view is to accommodate the (alleged) phenomenon of *causation by absence*. For example, it would seem that my failure to set my alarm clock last night caused me to sleep until 9 this morning. But my failure to set my alarm clock last night does not seem to be a candidate for an *event* – either intuitively or according to Lewis's account of events, since to characterise an 'event' as *failing* to do something is to characterise it in extrinsic terms (in terms of what is *not* going on in a given spatiotemporal region, rather than in terms of what *is* going on). By contrast, it is uncontroversially a *fact* that I failed to set the alarm clock, and so it can count uncontroversially as a causal truth that I slept until nine *because* I failed to set my alarm clock. I say more about causation by absence below.

Some authors have argued that *objects* themselves can be causes and effects. For example, one might claim that the bomb itself caused the explosion, and not merely the various events that involved the bomb in some way (my manufacturing and placing the bomb, lighting the fuse, and so on). In particular, some authors have claimed that *agents* (though not objects, such as bombs, more generally) are – or at least *could* be – causes of their own actions, in a way that does not simply reduce to their mental states (which we might think of as a species of event) causing their actions (see O'Connor 1995). Others hold that a unified account of causal relata – that is, an account that recognises just one ontological category as a suitable relatum for causation – is preferable (Menzies 1989; see Schaffer 2008, §1 for an overview of the debate about causal relata).

Causal relevance

When one event causes another, only some of its features – or some of the properties of the objects involved in the event – will be causally relevant to the effect. For example, suppose I sink the black ball while wearing a blue jumper and humming to myself. The angle at which I hit the cue ball and the amount of force the cue stick imparted to it are both relevant to the black going in the pocket (*e*), while neither the colour of my jumper nor my humming are relevant. To put matters counterfactually, *e* would still have occurred had I not been wearing a blue jumper, or had I not been humming to myself. By contrast, it wouldn't have occurred if I had hit the cue ball at a (very) different angle or much more gently. How should we account for the causal relevance or irrelevance of properties?

Lewis's answer is in effect already built into his distinction between the essential and accidental features of events. *e* would still have occurred had the event that is essentially my hitting the cue ball while humming (*c1*) not occurred, since (arguably) the closest possible world where *c1* does not occur is a world where I hit the cue ball just as I did but without humming at the same time; while *e* would *not* still have occurred had the event that is essentially my hitting the cue ball at the angle I did (*c2*) not occurred, since the closest possible world where *c2* does not occur is a world where I hit the cue ball at a different angle and the black therefore misses the pocket. Lewis's revised (2000) account of causation complicates things rather, however; and we might also wonder whether his theory of events really gets the right answers all of the time. For example, is the closest possible world where *c3*, my hitting the cue ball while wearing a blue jumper, a possible world where I'm not wearing the blue jumper (in which case *e* doesn't counterfactually depend on *c3*), or is it a world where I don't hit the ball at all (in which case *e* counterfactually depend on *c3*)?

Others have argued that consideration of causal relevance reveals that causation is a *contrastive* relation: rather than simply a binary relation between two events *c* and *e*, *c* causes *e relative* to a given contrast class (see e.g. Maslen 2004; Schaffer 2005). Consider my sinking of the black again. Did my hitting the cue ball with force *f* (*c*) cause the black to go in the pocket (*e*)? The contrastivist says that the answer is: it depends what the contrast case is. If I'd hit it with just a tiny bit less force (*f1*, let's say), then I still would have sunk the black; on the other hand, if I'd hit it with a *lot* less force (*f2*), the black wouldn't have made it all the way to the pocket.

So relative to the contrast case *f1*, *c* was not a cause of *e*; relative to *f2*, *c was* a cause of *e*.

One issue where the issue of causal relevance is crucially important is in the 'causal exclusion problem'. Many properties we ascribe to objects are 'multiply realisable', or stand in 'determinable' relations to 'determinate' properties. For example, there are many ways for an object to be *red* (redness is a 'determinable' property): it can be scarlet, maroon, pillar-box red, and so on (think of these as the 'determinate' properties). Similarly, suppose we adopt a 'functionalist' account of mental properties (See §4 above). Then it might well be that mental properties are multiply realisable; for example it might be that many different brain states, aside from the firing of C-fibres, can play the pain role, or that many different brain states can equally well 'realise' the property of believing that it is raining. In fact, the vast majority of properties we normally ascribe to objects are probably determinable or multiply realisable properties: there are many ways to be square, or cold, or wet, or whatever.

This raises what is sometimes known as the 'exclusion problem' – a generalised version of the 'problem of mental causation' (see Robb and Heil 2009), which in effect is the exclusion problem as applied to the particular case of mental states. Very roughly, the problem is how to account for the causal relevance of mental properties – or of determinable or functional properties more generally. Suppose (as seems plausible) that in principle there exists a complete causal explanation of any given event that just appeals to realiser or determinate properties. For example, if we knew enough neuroscience (say), we would be able to fully explain why Jane just pressed the button on the vending machine just by appealing to neurological or, more broadly, physical properties of Jane and her environment: her brain was in a certain state, which resulted in certain signals being sent to her muscles, which caused her finger to apply pressure to the button. But we also want to say that Jane's *wanting a cold drink* – the *mental* state she was in - was causally relevant to her pressing the button. But how can that be, when in principle we don't need to appeal to her mental state at all in giving a complete causal explanation of her pressing the button? To put it another way, the causal relevance of Jane's *neurological* state seems to exclude the possibility that her *mental* state is *also* causally relevant. Of course, in general more than one property is relevant to any given effect; both the force and the angle of my snooker shot were causally relevant to sinking the black, for example. But in this kind

of case we would not have a *complete* causal explanation if we omitted one of those properties: I can't *fully* explain why the black went in the pocket without mentioning *both* the force and the angle of the shot. In the case of Jane's pressing the button, however, it seems we can make do with just the physical property; so it's unclear that we're entitled to think that mental properties (and determinable and functional properties more generally) do any causal work.

This is worrying, since in the absence of a good response to the exclusion problem we are saddled with the conclusion that determinable and functional properties are merely epiphenomenal. Various attempts have been made to resolve the exclusion problem; see for example Yablo 1992 and Bennett 2003.

One concept of causation or two?

Let's return to causation by absence, mentioned above. This is an instance of what Jonathan Schaffer (2000) calls 'causation by disconnection': putative cases of causation that involve no *process* connecting the cause to the effect. Another kind of (possible) causation by disconnection is the spell-casting case described earlier. Still another is'double prevention'. Suppose Jane sees the assassin aim his gun at the President, wrestles him to the ground and deprives him of the gun (*c*), thereby preventing him from shooting, which would in turn have prevented the President from arriving unharmed at his intended destination (*e*). Is *c* a cause of *e*? Certainly *e* counterfactually depends on *c* – if Jane hadn't stopped the assassin from shooting, the President would not have arrived unharmed at his destination. And the case has many other standard hallmarks of causation: for example, Jane wanted (let's suppose) to save the President's life, and her action was an excellent means for realising that end. On the other hand, there is no causal process that connects the two events: here is Jane wrestling the assassin to the ground, and there is the President, some distance away, oblivious to what is going on on the grassy knoll.

When we consider the kinds of theory described in §4 above, they fall into three different camps when it comes to causation by disconnection. Process and mechanistic theories do not recognise causation by disconnection: for example there is no transmission of any conserved quantity from *c* to *e* in the above case, and nor is there any mechanism connecting the two. (For this reason Dowe argues that 'causation' by disconnection is not really causation at all, but what he calls '*quasi*-causation' (Dowe 2001).) Probabilistic theories, by contrast, have no problem with

causation by disconnection, because all that is important for causation is the statistical relationship between causes and effects – and such a relationship can clearly exist in the absence of any kind of connecting mechanism. (Wrestling the gunman to the ground significantly increases the probability that the intended victim will survive, for example.) When it comes to counterfactual theories of causation, one might go either way. On the one hand, counterfactual dependence just by itself does not require any kind of process connecting cause and effect – as we just saw in the Jane-and-the-assassin case. On the other hand, if we retain the idea that causes and effects must be Lewisian events, then at least some alleged cases of causation by disconnection – in particular, cases where the cause or the effect is an absence – will turn out not to be cases of causation at all, since, as we saw earlier, absences are not Lewisian events. So we might retain our commitment to the idea that causation is a relation between Lewisian events and attempt to explain away alleged cases of causation by absence (see Beebee 2004), or else – and this is Lewis's favoured approach – we might abandon the idea that causation is a relation between Lewisian events (Lewis 2004).

Ned Hall (2004) has argued that the moral we should draw is that there are in fact two distinct concepts of causation: what he calls 'production' and 'dependence'. Dependence is characterised in counterfactual terms and production is a relation that is transitive, intrinsic and 'local': it is constituted by the existence of a spatiotemporally continuous sequence of intermediate events. In most ordinary cases of causation the two concepts both apply, and so the existence of two distinct concepts is not immediately obvious. (When my hitting the cue ball (*c*) causes the red to go in the pocket (*e*), there is both a process leading from one to the other (production) and counterfactual dependence of *e* on *c*.) However, the two can come apart. The Jane-and-the-assassin case is one of dependence without production, and there are also cases, such as pre-emption cases (see §4 above), where we have production without dependence. (Hall's proposal is discussed in e.g. Williamson 2006 and Godfrey-Smith 2009.)

Peter van Inwagen describes 'cause' as a 'horrible little word', adding: 'Causation is a morass in which I for one refuse to set foot. Or not unless I am pushed' (1983, 65). Unfortunately for van Inwagen, the centrality of causation for our understanding the world *does* require us to set foot in the morass. Fortunately, many philosophers have done so. While there is no broad consensus on any of the issues

discussed in this chapter, it is surely true that causation is less of a morass than it once was.

Bibliography

Anscombe, G. E. M. 1971. *Causality and Determination: An Inaugural Lecture*. Cambridge: Cambridge University Press. Reprinted in E. Sosa and M. Tooley (eds), *Causation*, Oxford: Oxford University Press, 1993.

Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

Aronson, J. L. 1971. 'On the grammar of "cause"', *Synthese,* 22: 414-30.

Beebee, H. 2003. 'Seeing causing', *Proceedings of the Aristotelian Society,* 103, 257-80.

-- 2004. 'Causing and nothingness', in Collins, Hall & Paul 2004.

-- 2006. *Hume on Causation.* Abingdon: Routledge.

-- 2009. 'Causation and observation', in Beebee, Hitchcock & Menzies 2009.

--, C. Hitchcock & P. Menzies (eds). 2009. *The Oxford Handbook of Causation.* Oxford: Oxford University Press.

Bennett, J. 1988. *Facts and their Names.* Indianapolis: Hackett.

Bennett, K. 2003. 'Why the exclusion problem seems intractable, and how, just maybe, to tract it', *Nous*, 37: 471-97.

Collins, J., E. J. Hall & L. A. Paul (eds). 2004. *Causation and Counterfactuals*. Cambridge, MA: MIT Press.

Davidson, D. 1967. 'Causal relations', *The Journal of Philosophy*, 64: 691-703.

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

-- 2001. 'A counterfactual theory of prevention and "causation" by omission', *Australasian Journal of Philosophy,* 79: 216-26.

-- 2008. 'Causal processes', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*. URL = <http://plato.stanford.edu/archives/fall2008/entries/causation-process/>.

Dupré, J. 1984. 'Probabilistic causality emancipated', in P. French, T. Uehling, Jr., & H. Wettstein (eds), *Midwest Studies in Philosophy IX*, Minneapolis: University of Minnesota Press, 169–75.

-- 1990. 'Probabilistic causality: a rejoinder to Ellery Eells', *Philosophy of Science*, 57: 690-98.

Eells, E.1987. ‘Probabilistic causality: reply to John Dupré’, *Philosophy of Science,* 54: 105-14.

-- 1991. *Probabilistic Causality*. Cambridge: Cambridge University Press.

Fair, D. 1979. ‘Causation and the flow of energy’, *Erkenntnis*, 14: 219-50.

Glennan, S. 1996. ‘Mechanisms and the nature of causation’, *Erkenntnis*, 44: 49-71.

Godfrey-Smith, P. 2009. ‘Causal pluralism’, in Beebee, Hitchcock & Menzies 2009.

Goldman, A. 1967. ‘A causal theory of knowing’, *the Journal of Philosophy,* 64: 357-72.

Grice, H. P. 1961. ‘The causal theory of perception’, *Proceedings of the Aristotelian Society,* Supp. Vol. 35: 121-53.

Hall, N. 2004. ‘Two concepts of causation’, in Collins, Hall & Paul 2004.

Heathcote, A. & D. M. Armstrong. 1991. ‘Causes and laws’, *Nous,* 25: 63-73.

Hitchcock, C. 2011. ‘Probabilistic causation’, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*. URL = <http://plato.stanford.edu/archives/win2011/entries/causation-probabilistic/>.

Hume, D. 1739-40. *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd edition, revised and ed. P.H. Nidditch, Oxford: Clarendon Press (1978).

-- 1748/1751. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, 3rd edition, revised and ed. P.H. Nidditch, Oxford: Clarendon Press (1975).

Kim, J. 1973. ‘Causation, nomic subsumption, and the concept of event’, in *The Journal of Philosophy,* 70: 217-36.

Levin, J. 2010. ‘Functionalism’, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*. URL = <http://plato.stanford.edu/archives/sum2010/entries/functionalism/>.

Lewis, D. K. 1973. ‘Causation’, *The Journal of Philosophy*, 70: 556–67. Reprinted with postscripts in Lewis 1986.

-- 1973a. *Counterfactuals*. Oxford: Blackwell.

-- 1979. ‘Counterfactual dependence and time's arrow’, *Noûs*, 13: 455-76. Reprinted in Lewis 1986.

-- 1981. ‘Causal decision theory’, *Australasian Journal of Philosophy*, 59: 5–30.

-- 1986. *Philosophical Papers, Vol. II.* Oxford: Blackwell.

-- 1986a. ‘Events’, in Lewis 1986.

-- 2000. 'Causation as influence', *The Journal of Philosophy,* 97: 182-97. Reprinted in Collins, Hall & Paul 2004.

-- 2004. 'Void and object', in Collins, Hall & Paul 2004.

Machamer, P., L. Darden & C. Craver. 2000. 'Thinking about mechanisms', *Philosophy of Science,* 67: 1-25.

Mackie, J. L. 1965. 'Causes and conditions', *American Philosophical Quarterly*, 2: 245-64. Reprinted in E. Sosa & M. Tooley (eds.), *Causation*, Oxford: Oxford University Press, 1993.

Maslen, C. 2004. 'Causation, contrasts, and the nontransitivity of causation, in Collins, Hall & Paul 2004.

Mellor, D. H. 1995. *The Facts of Causation.* New York: Routledge.

Menzies, P. 1989. 'A unified account of causal relata', *Australasian Journal of Philosophy,* 67: 59-83.

-- 1996. 'Probabilistic causation and the pre-emption problem', *Mind,* 105: 85-117.

-- 1998. 'How justified are the Humean doubts about intrinsic causal links?', *Communication and Cognition*, 31: 339-64.

-- and H. Price. 1993. 'Causation as a secondary quality', *British Journal for the Philosophy of Science,* 44: 187-203.

Noordhof, P. 1999. 'Probabilistic causation, preemption and counterfactuals', *Mind*, 108: 95-125.

O'Connor, T. 1995. 'Agent causation', in T. O'Connor (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press.

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

Price, H. 2007. 'Causal perspectivalism', in H. Price & R. Corry (eds), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press.

Quine, W. V. 1963. 'Reference and Modality', in his *From a Logical Point of View*. New York: Harper & Row.

Ramsey, F. P. 1929. 'General propositions and causality', in D. H. Mellor (ed.), *F. P. Ramsey: Philosophical Papers,* Cambridge: Cambridge University Press, 1990.

Reichenbach, H. 1928. *The Philosophy of Space and Time*. 1958 edn. trans. M. Reichenbach & J. Freund, New York: Dover.

Robb, D. & J. Heil. 2009. 'Mental causation', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*, URL = <http://plato.stanford.edu/archives/sum2009/entries/mental-causation/>.

Russell, B. 1912-13. 'On the notion of cause', *Proceedings of the Aristotelian Society*, 13: 1-26.

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

Schaffer, J. 2000. 'Causation by disconnection', *Philosophy of Science,* 67: 285-300.

-- 2000a. 'Trumping pre-emption', *The Journal of Philosophy*, 97: 165-181.

-- 2001. 'Causation, influence, and effluence', *Analysis,* 61: 11-19.

-- 2005. 'Contrastive causation'. *The Philosophical Review,* 114: 297-328.

-- 2008. 'The metaphysics of causation', in E. N. Zalta (ed.),*The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*. URL = <http://plato.stanford.edu/archives/fall2008/entries/causation-metaphysics/>.

Suppes, P. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Publishing Company.

van Inwagen, P. 1983. *An Essay on Free Will.* Oxford: Clarendon Press.

Williamson, J. 2006. 'Causal pluralism versus epistemic causality', *Philosophica,* 77: 69-96.

-- 2011. 'Mechanistic theories of causality', *Philosophy Compass*, 6: 421-47.

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

-- 2008. 'Causation and manipulability', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2008 Edition)*. URL = <http://plato.stanford.edu/archives/win2008/entries/causation-mani/>.

Yablo, S. 1992. 'Mental causation', *The Philosophical Review,* 101: 245-80.

Annotated bibliography (these all appear above as well)

Lewis, D. 2000. 'Causation as influence', *The Journal of Philosophy,* 97: 182-97. Reprinted in Collins, J., N. Hall & L. A. Paul (eds). 2004. *Causation and Counterfactuals*. Cambridge, MA: MIT Press.

This is David Lewis's final attempt to construct a viable counterfactual analysis of causation. It defines causation not in terms of 'whether-whether' dependence (so that whether or not the effect occurs counterfactually depends on whether the cause occurs), but rather in terms of 'influence' or the extent to which 'alterations' of the effect counterfactually depend on 'alterations' of the cause. Thus *c influences e* to the extent that whether, how and when *e* occurs counterfactually depends on whether, how and when *c* occurs.

H. Beebee, C. Hitchcock & P. Menzies (eds). 2009. *The Oxford Handbook of Causation.* Oxford: Oxford University Press.

This is a comprehensive collection aiming to summarise both the history of the causation debate and the current state of the many contemporary debates, including the role of causation in philosophical theories more widely and in other disciplines.

Hall, N. 2004. 'Two concepts of causation', in Collins, Hall & Paul 2004.

This paper puts forward the influential thesis that there are two concepts of causation rather than one: a 'dependence' concept and a 'production' concept. A large part of its interest rests on its diagnosis of, and suggested solution to, many of the issues surrounding the analysis of causation that have proved intractable in recent years.

Anscombe, G. E. M. 1971. *Causality and Determination: An Inaugural Lecture*. Cambridge: Cambridge University Press. Reprinted in E. Sosa and M. Tooley (eds), *Causation*, Oxford: Oxford University Press, 1993.

Anscombe's classic lecture aims to sound the death knell on Hume's influential claim that causation is to be understood in terms of regularity.