1 Introduction

We humans take a great interest in causation. Causal knowledge helps us to understand, predict, and influence the world around us. A baby quickly comes to realise that pushing a button on her toy causes it to play a song; an adult exploits her knowledge of the effects of chamomile to sooth the baby's teething gums. Because of the close relation between causation, understanding, prediction, and control, natural and social scientists devote significant time and resources to investigating causal questions. Among other things, they ask or have asked: What are the causes of cancer? Of climate change? Of anomalies in the orbit of Uranus? What caused the extinction of the dinosaurs? The First World War? The 2007-8 financial crisis? Trump's election? The Covid-19 outbreak? What causes mental health problems? Crime? Price inflation?

Causation is of importance to psychology: if we want to understand how humans learn and reason, we need to understand their capacity for *causal* learning and reasoning. It's of importance in AI: if we want computers and robots to learn as well as (or better than!) humans, and to manipulate the world as (or more!) effectively, we need to programme them to be able to acquire causal knowledge and to use it.

Causation is also closely tied to questions of moral and legal responsibility. In a landmark UK legal case, the wife of the late Arthur Fairchild successfully sued Glenhaven Funeral Services[1] over her husband's death from mesothelioma – a type of lung cancer caused by asbestos exposure. In order to establish the company's responsibility for Fairchild's fatal illness, it was of course necessary to establish that there was a causal link between something they'd done – namely negligently expose Fairchild to elevated levels of asbestos – and the illness itself.

Despite its ubiquity and importance, it's surprisingly difficult to say exactly what causation is. Difficult questions about the fundamental nature of the world – especially those that don't readily admit of empirical resolution – naturally attract the attention of philosophers. But causation isn't only of intrinsic philosophical interest. Greater theoretical clarity on its nature has had significant payoffs in the sciences and in law. And, close to home for philosophers, it has payoffs in virtue of the fact that causation plays a role in key theories of a variety of philosophically interesting phenomena including (but not limited to) reference, perception, decision, knowledge, inference, action, and explanation.

It shouldn't be thought that work on the theory of causation is the exclusive preserve of philosophers. Much important theoretical work has been done by

[1] *Fairchild v Glenhaven Funeral Services Ltd* ([2002] UKHL 22; [2003] 1 AC 32).

CAMBRIDGE

Cambridge University Press
978-1-108-70663-6 — Causation
Luke Fenton-Glynn
Excerpt
More Information

computer scientists, economists, statisticians, legal scholars, psychologists and others – reflecting the broad, interdisciplinary importance of a better understanding of causation. Fortunately, these days, there's significant interaction between theorists in these various disciplines, which has enriched our collective understanding of causation. A prime example of the payoffs of this cross-disciplinary interaction is the theory of causal modelling that we'll examine as part of Sections 4 and 5.

This Element examines some of the progress that's been made in understanding the nature of causation as well as some of the unresolved challenges. Since the literature on causation is large, this Element is, of necessity, a selective introduction. It focuses on three broad traditions within the theory of causation: the regularity, counterfactual, and probabilistic approaches. Perhaps the most contentious omissions are the process approach – which seeks to analyse causation in terms of causal processes with the latter understood, on the most promising such account, as the world-lines of objects that possess conserved quantities (Dowe 1992, 2000; Salmon 1994, 1997) – and the New Mechanist approach – exemplified by Machamer et al. (2000) and Glennan (2017), among many others. I've made this choice, not because I don't think that understanding causal processes and mechanisms is of vital importance (I do!), but because I don't think that these are causal bedrock: I think there are relations of causation that are more fundamental than the notion of a causal process or mechanism and that an adequate understanding of processes and mechanisms will require an adequate understanding of these more fundamental causal relations.[2] Regularity, counterfactual, and probabilistic approaches are attempts to understand these fundamental causal relations.

This view is contentious as process theorists and some New Mechanists think that the fundamental causal relation(s) *can* be understood in terms of processes or mechanisms. For example, Dowe (2000, 90) seeks to define causal interactions in terms of processes, while Glennan (1996) suggests that causation – at least outside the domain of microphysics – might be analysed in terms of mechanisms. However, process theories have been plagued by the problem of distinguishing genuine causal processes from world-lines that don't correspond to processes without falling back on an appeal to some more basic causal relation, perhaps understood in terms of counterfactual dependence (see, e.g., Hitchcock 1995, 2009; Choi 2002).

As regards mechanistic approaches, there's certainly no consensus among New Mechanists that the notion of mechanism is prior to that of causation. As

[2] The notion that there might be more than one fundamental causal relation is taken up in Section 3.4.2.

Cambridge University Press
978-1-108-70663-6 — Causation
Luke Fenton-Glynn
Excerpt
More Information

Craver and Tabery (2019) note, ‘Mechanists have disagreed with one another about how to understand the cause in causal mechanism. … Four ways of unpacking [it] have been discussed: conserved quantity accounts, mechanistic accounts, activities accounts, and counterfactual accounts’.

I’ve already said that it’s doubtful that the conserved quantity approach can yield an understanding of the fundamental causal relation(s). The activities approach, on the other hand, is a primitivist approach (see Craver and Tabery 2019), whereas we’ll be examining accounts that seek a deeper understanding of causation. Meanwhile, the counterfactual approach is one that we’ll be exploring in Section 4. Finally, the mechanistic account – as advocated by Glennan (1996) – is regressive. The proposal is that causal connections that may seem basic at (say) the biological level can be understood in terms of mechanisms at the chemical level, and those at the chemical level in terms of mechanisms at the physical level. There’s thus a hierarchy of mechanisms. The concern, though, is that this hierarchy bottoms out at the level of fundamental physics at which level we have causings that can’t be mechanistically understood. Again, this favours the view that there are fundamental causal relations in terms of which mechanisms can ultimately be understood. The regularity, counterfactual, and probabilistic approaches seem the most promising approaches to understanding these basic causal relations.

2 Regularity Theories of Causation

2.1 Hume

Though Western theorising about causation dates back at least to Aristotle (*Physics* 195 a 4–14; *Metaphysics* V.2), David Hume (1739, 1748) is rightly considered the father of the modern tradition of attempts to understand that relation. Hume is standardly interpreted as advocating a *regularity theory* of causation.

Specifically, according to Hume (1739 I.iii.2), causes occur temporally prior to their effects, and are either contiguous with them in space and time or else connected to them by a contiguous ‘chain’ of causation. For example, a person may break a window by throwing a rock at it despite the fact that the throw occurs a short interval of time before and a few metres away from the breaking because, once the rock is thrown, it traces a continuous trajectory until it hits and shatters the window. The window’s breaking is caused by the throw via this ‘chain’ (there’s no ‘action at a distance’), with the position and momentum of the rock at each stage on its trajectory being caused by its previous positions and momenta and by the throw itself, and with the window’s shattering being caused by the prior states of the rock all the way back until we get to the throw itself.

But we don't have a case of causation just any time an event occurs prior to and contiguously with another. Towards the end of the movie *Saving Private Ryan*, in a defiant last stand, Captain Miller repeatedly fires his pistol at a German Tiger tank. The bullets are of course completely incapable of piercing the tank's armour. Down to his last bullet, Miller points his gun and shoots at the tank at the very moment the tank is blown to pieces by a bomb dropped by a US P-51 aircraft. The impact of Miller's bullet is immediately prior to, and contiguous with, the explosion of the tank. Yet it's the bomb and not the bullet that causes the tank to explode.

Fortunately, Hume's account doesn't imply that the bullet impact was a cause. That's because, in addition to priority and contiguity, Hume adds a third requirement: *constant conjunction*. For Hume, for an event *c* to be a cause of an event *e*, it must be the case that events like *c* are always followed contiguously by events like *e*.[3] This criterion excludes the impact of Miller's bullet from counting as a cause of the tank's explosion. That's because events like the former aren't always followed by events like the latter. Indeed, Miller had already fired his gun at the tank five times prior to firing his last bullet: the impact of none of these previous five bullets was followed contiguously by an explosion. So Hume's analysis yields the correct verdict about this case.

Although Hume doesn't say this, it's tempting to think that not any old constant conjunction can ground a causal relation, but rather one might wish to require that the constant conjunction be entailed by the laws of nature. This avoids problems such as the following. Suppose there exists an extremely rare isotope, call it 'unobtanium-352'. Only one atom of this isotope ever exists. Suppose this atom happened to decay on the afternoon of November 19, 1863, immediately before Lincoln delivered the Gettysburg Address and contiguously with it. Now, for any type *T* of event of which the Gettysburg Address is a member ('famous speeches', say), it's true that all cases of unobtanium-352 decay are followed by events of type *T*. Nevertheless, it clearly doesn't follow that the decay of the unobtanium-352 atom was a cause of the Gettysburg Address.[4] A sophisticated regularity theory can avoid this

[3] We could, on Hume's behalf, distinguish *direct* from *indirect* causation, with *c* counting as a direct cause of *e* iff *c* is prior to, and contiguous with, *e* and events like *c* are always followed contiguously by events like *e*. *Indirect* causation would then be understood in terms of chains of (i.e. ordered sequences of events that stand in relations of) direct causation.

[4] Note that even if we take 'constant conjunction' to require a multiple instances, we're still liable to get 'accidental' constant conjunctions that aren't apt to underwrite causal relations (see Armstrong 1983, 15–17).

conclusion by pointing out that the fact that all instances of unobtanium-352 decay are followed by events of type T isn't entailed by the laws of nature (rather, it's an instance of an accidental regularity).[5]

2.2 Mill

An apparent problem with Hume's account is that sometimes we seem to have causation without constant conjunction. John Stuart Mill pointed this out in making the following observation:

> It is seldom, if ever, between a consequent and a single antecedent that … invariable sequence subsists. It is usually between a consequent and the sum of several antecedents … . In such cases it is very common to single out only one of the antecedents under the denomination of Cause, calling the others merely Conditions. Thus, if a person eats of a particular dish, and dies in consequence … people would be apt to say that eating of that dish was the cause of death. There needs not, however, be any invariable connexion between eating of the dish and death; but there certainly is, among the circumstances which took place, some combination or other on which death is invariably consequent: as, for instance, the act of eating of the dish, combined with a particular bodily constitution … . The real Cause, is the whole of these antecedents … . (Mill 1843, III.v.3)

Mill's idea, then, is that constant conjunction ('invariable sequence') rarely obtains between a single earlier event ('antecedent') and a single later event ('consequent'). Taking Mill's example, it could easily happen that someone dies from eating a particular dish (because, say, it contains an allergen) but that others who eat it survive. Mill thinks that those who die have features that distinguish them from those who don't. For instance, they may have a severe allergy to an allergen in the dish. In this case, the constant conjunction

[5] A couple of comments are worth making regarding this proposed appeal to laws of nature. *First*, whether one regards it as marking a departure from a pure regularity theory of causation will depend upon one's preferred metaphysics of laws. It will not be a departure if one adopts a *regularity theory of laws*. Whilst sophisticated regularity theories of laws – such as the Best System Analysis (Lewis 1994) – regard laws as regularities, they don't count just any old regularity as a law of nature. Thus, for instance, it's to be hoped that they wouldn't count an unobtainium-325 decay/famous speech regularity as a law.

Second, for the appeal to laws of nature to be satisfactory, it will presumably be necessary that a wide range of regularities outside the domain of fundamental physics (e.g. 'aspirin consumption is followed by pain relief') are entailed by laws of nature. That's because it's clear that our causal claims extend to such domains. There's an extensive philosophical literature on the status of generalisations outside of fundamental physics. For overviews of important aspects of this literature, see Cat (2017) and Reutlinger et al. (2019). Thanks to an anonymous referee for encouraging me to say something about both of the foregoing points.

is between eating the dish *and (e.g.) having a severe peanut allergy* (as Mill puts it: having 'a particular bodily constitution') and death.

Mill thinks that, properly speaking, in such a case it's the combination of the severe peanut allergy with the eating of the dish that's the cause of death. However he observes elsewhere (Mill 1843, III.v.3) that, in ordinary talk, we often single out just one of the factors in such a combination as the cause and regard the others as mere 'conditions'. Specifically, he claims that we're inclined to pick out '*events*' or '*changes*' like the eating of the dish as causes and treat long-standing '*states*' like the possessing of the peanut allergy as mere '*conditions*'.

2.3 Hart and Honoré on the Cause/Condition Distinction

As noted in the Introduction, the study of causation is a truly interdisciplinary endeavour and the profound contributions of the legal scholars H. L. A. Hart and Tony Honoré to our understanding of it is the perfect illustration of this. One of their contributions was to further investigate the relationship between those factors that we pick out as 'causes' and those that we label mere 'conditions'. According to them, this distinction is one 'to which common sense adheres in face of the demonstration that cause and conditions are "equally necessary" if the effect is to follow' (Hart and Honoré 1959, 33). In giving an account of the grounds on which we make such distinctions, they emphasise the close connection between causation and explanation.

Hart and Honoré (1959, 35) point out that the sorts of effect that tend to pique our interest – and lead us to seek causes to explain – are abnormal events. In Mill's example, this was the sudden death of a person. Previous examples that we've used include the shattering of a window and the explosion of a tank. Hart and Honoré consider the example of the outbreak of a fire. They point out that, in order to explain an abnormal event, some abnormal factor typically needs to be cited. For example, in order to explain why a building was destroyed by fire, emphasising the presence of oxygen or flammable material wouldn't typically be that helpful, whereas pointing out that a lit cigarette was dropped would be much more so. The reason is that oxygen and flammable material were *always* present, but what we need for an explanation is to know what *made the difference* between this occasion, on which the building burned down, and all the previous times when it didn't (Hart and Honoré 1959, 35). It's the abnormal factor, the dropping of the cigarette, that's the difference-maker and hence we're liable to call it the 'cause' while treating the others as 'mere conditions' (Hart and Honoré 1959, 35).

Normal factors will often be what Mill described as 'states' – ongoing or permanent factors – while abnormal factors will often be 'changes' or 'events'. For instance, the presence of oxygen and flammable material are typically relatively permanent states, while the dropping of a cigarette is a 'change'. But Hart and Honoré show that things aren't that simple. They give the following example:

> If a fire breaks out in a laboratory or in a factory, where special precautions are taken to exclude oxygen during part of an experiment or manufacturing process, since the success of this depends on safety from fire, there would be no absurdity at all in *such* a case in saying that the presence of oxygen was the cause of the fire. The exclusion of oxygen in such a case, and not its presence, is part of the normal functioning of the laboratory or factory … . (Hart and Honoré 1959, 35)

Suppose that the manufacturing process involves the frequent production of sparks. Then, it seems, we might pick out the presence of oxygen as a cause even if the leak occurred some time before a spark was produced and the fire started. If that's right, then we would appear to have a case where the relatively longstanding state (the presence of oxygen) is picked out as the cause, while the change (the spark) might be treated as a mere condition. This would suggest that the distinction we draw between causes and conditions tracks the abnormal/normal distinction rather than the event/state distinction where these two come apart.

2.4 Mackie

As we saw, although Mill acknowledges that in ordinary talk we distinguish between causes and conditions, he thinks that the 'real Cause' is the combination of all the factors needed to bring about the effect. We also saw that Hart and Honoré seek to account for why we distinguish between causes and conditions in terms of our explanatory interests, noting that these various factors may be equally necessary for the effect. Especially if we think that the event/state distinction does not always track the cause/condition distinction (as Hart and Honoré's example suggests), we might conclude that the cause/condition distinction isn't really a metaphysical distinction at all but rather one that's to be accounted for by a suitable pragmatics of causal talk. It's tempting to impute this view to David Lewis when he says:

> We sometimes single out one among all the causes of some event and call it 'the' cause, as if there were no others. Or we single out a few as the 'causes,' calling the rest mere 'causal factors' or 'causal conditions.' … We may select the abnormal or extraordinary causes, or those under human control, or those

> we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). (Lewis 1973a, 558–9)

We'll return to Lewis's account of causation in Section 4. For now it's worth noting that, even if we agree that there's no deep metaphysical distinction between causes and conditions (or indeed between the factor that we might on some occasion be inclined to pick out as 'the' cause of some effect, and the rest of that effect's causes), we needn't follow Mill in taking the conjunction of all the factors requisite to produce an effect as the real cause. An alternative approach – as suggested in the passage from Lewis – is that we allow that each of the factors in question counts as *a* cause, so that a given effect may have a plurality of causes.

The latter is the approach of Mackie, whose well-known account of causation perhaps marked the zenith of the regularity approach.[6] Mackie's account can be introduced by means of his own example:

> Suppose that a fire has broken out in a certain house …. Experts … conclude that it was caused by an electrical short-circuit at a certain place. … Clearly the experts are not saying that the short-circuit was a necessary condition for this house's catching fire at this time; they know perfectly well that … the overturning of a lighted oil stove, or any one of a number of other things might, if it had occurred, have set the house on fire. Equally, they are not saying that the short-circuit was a sufficient condition for this house's catching fire; for if the short-circuit had occurred, but there had been no inflammable material nearby, the fire would not have broken out …. In what sense, then, is it said to have caused the fire? At least part of the answer is that there is a set of conditions … including the presence of inflammable material, the absence of a suitably placed sprinkler … which combined with the short-circuit constituted a complex condition that was sufficient for the house's catching fire – sufficient, but not necessary, for the fire could have started in other ways. Also, of *this* complex condition, the short-circuit was an indispensable part: the other parts of this condition, conjoined with one another in the absence of the short-circuit, would not have produced the fire. … In this case, then, the so-called cause is … an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result. … [L]et us call such a condition (from the initial letters of the words italicized above), an inus condition. (Mackie 1965, 245)

[6] This isn't to denigrate subsequent accounts within the regularity tradition, including the excellent contributions of Strevens (2007) and Baumgartner (2013). But these can be regarded as attempts to revive the tradition after a prolonged period in which it has been out of favour. It's fair to say that it currently remains a minority approach.

Clearly Mackie's account makes heavy use of the notions of *necessary* and *sufficient* conditions. The reason it's usually classed as a regularity theory is that necessary and sufficient conditions can themselves be understood in terms of regularities.[7] For instance, we might say that the set of conditions comprising the short-circuit, the presence of inflammable material, the presence of oxygen, the absence of a sprinkler, etc. is *sufficient* for the fire iff whenever such a constellation of factors co-occurs, a fire always ensues. Likewise, we might say that the short-circuit is a *necessary* (or non-redundant) element of this set of conditions iff it's not the case that whenever the remaining conditions in the set co-occur then fire ensues.[8]

With the notions of 'necessity' and 'sufficiency' so interpreted, we can view Hume as taking causes to be sufficient for their effects (and possibly necessary too, if 'constant conjunction' is taken to cut both ways, so that not only does a cause never occur without its associated effect, but an effect never occurs without its associated cause), while Mill agrees that real causes are sufficient for their effects but takes them to typically be complex (e.g. the complex condition comprising the short-circuit, the presence of oxygen, the presence of inflammable material, and the absence of a sprinkler system, etc.). Mackie, on the other hand, allows non-redundant elements of such sets (e.g. the short-circuit) to count as causes.

2.5 Problems with the Regularity Theory

Perhaps the most important problems that have been raised against the regularity approach – and which have contributed to its decline in popularity in recent years – are (i) the problem of the direction of causation; (ii) the problem of probabilistic causes; (iii) the problem of common causes. We'll see in later sections that other theories of causation aren't immune to these problems, but they're often thought to be particularly intractable for the regularity approach.

2.5.1 The Direction of Causation

Causation is almost always supposed to be an asymmetric relation: if a causes b, then b doesn't also cause a. One might object that feedback loops provide a counterexample: depression causes someone to drink, but alcohol itself acts as a depressant. The rejoinder is that, while a person's depressed state at time t_1

[7] Though, drawing upon the point made in Section 2.1, we may wish to understand them in terms of *lawfully entailed* regularities.

[8] Actually, Mackie himself proposes that necessity and sufficiency only indirectly be understood in terms of regularities: specifically, he proposes to interpret necessity and sufficiency in terms of counterfactuals, with the counterfactuals understood in terms of regularities (Mackie 1965, 253–5). So one might take Mackie's account to be a sort of hybrid regularity/counterfactual theory. Counterfactual theories will be examined in Section 4.

Cambridge University Press
978-1-108-70663-6 — Causation
Luke Fenton-Glynn
Excerpt
More Information

might cause her to drink at time t_2, which in turn causes her depressed state at time t_3, what we'd never have is a person's depressed state at t_1 causing her to drink at t_2 with her drinking at t_2 causing her depression at t_1. The point is that, where a and b are particular events or states that obtain at specific times, it can't be the case both that a causes b and that b causes a. This point is sometimes put by saying that *token causation* (that is, the causal relation between token – i.e. particular, dated – events or states) is asymmetric. We'll have more to say about token causation – and its distinction from what's sometimes known as *type* causation – in Section 3.3.

One might wonder whether it's really entirely impossible that a should be a token cause of b and b a token cause of a. The General Theory of Relativity (GTR) allows for the possibility of what are known as 'closed time-like curves' – paths in space-time of positive distance that lead from a given space-time point p back to p that could be traversed without ever travelling at or above the speed of light. Suppose object o traverses such a path from p to p and goes through some point q on its way. Then the state of o at p (its position and velocity) is presumably a cause of its state at q and its state at q a cause of its state at p. But, whilst GTR allows that this is a possibility, we don't have reason to think that closed time-like curves in fact exist in our universe.

The trouble for regularity theorists is that they face a challenge in accounting for why we don't have bi-directional causation even in quoditian cases. That's because necessity and sufficiency are two sides of the same coin: if a is sufficient for b then b is necessary for a. Putting it in regularity-theoretic terms, if events like a are always accompanied by events like b, then there are no a-like events without b-like events. This means that a challenge arises for the accounts of Hume and Mill when we have causes that are both necessary and sufficient for their effects. It also turns out that often when a is an inus condition of b, then b is also an inus condition of a.

This difficulty isn't necessarily fatal to the regularity theory. What it shows is that some extra element needs to be added to the analysis to distinguish cause from effect. Hume takes this extra element to be *time*: he requires that the cause be earlier in time than the effect. One potential objection to Hume's approach is that it's not clear how to reconcile it with the possibility of closed causal loops of the sort seemingly allowed by GTR. Mackie was reluctant to rule out backwards-in-time causation, and so instead took the cause to be the event that becomes *fixed* first (Mackie 1980, ch. 7). While, ordinarily, earlier events become fixed before later ones (because becoming a past event is the usual way in which an event becomes fixed), Mackie thought that in special circumstances (of the sort that would allow for retro-causation) a later event might become fixed prior to an earlier one. Yet it's not clear that Mackie's proposal