Causation: Further Themes
Alison Fernandes
October 2018

Article Summary
Recent work in the philosophy of causation has explored a number of issues relating to the objectivity of causation, including the place of causation in metaphysics and science, its temporal asymmetry, and whether causation is context-sensitive.

Regarding the place of causation, some have argued that causation is fundamental, such that the most basic level of reality is structured in causal terms (Tooley). Others take causation to be a higher-level phenomenon, and either attempt to derive causal relations from more basic non-causal relations (Lewis), or accept that causal relations are irreducible (Woodward). Others still take a more skeptical approach, and question the need for giving a metaphysics of causation (Price). How one conceives of causation’s place has implications for both the relation between fundamental physics and higher-level sciences, as well as the relation between science and metaphysics.

Regarding the temporal asymmetry of causation, a characteristic feature of causation is that causes come prior in time to their effects. We might accept this temporal asymmetry as a ‘metaphysical primitive’—a basic feature of causation not to be further explained. But some, particularly those defending reductive accounts (Lewis, Loewer), have attempted to explain causation’s temporal asymmetry by relating it to other temporal asymmetries, such as those of evidence or entropy. While most accept that explanations of causation’s temporal asymmetry must appeal to special initial conditions, there is debate over what kind of initial conditions are required.

Finally, counterfactual accounts of causation seem able to recover much of what one might want from an account of causation. But it is unclear whether such accounts agree (or can be made to agree) with our intuitive causal judgements about core cases. Some remain skeptical that any counterfactual account of causation can succeed. Others have taken putative counterexamples to show that causation is context-sensitive, such that
broadly pragmatic concerns are relevant to whether causal claims are either true, or appropriate.

1. Fundamentality and Physics
Recent work in the philosophy of causation has explored the place of causation in metaphysics and science. Are causal relations part of the fundamental ontology of the world, such that the most basic level of reality has causal structure? Or is causation a higher-level phenomenon, which appears when we consider what derives from this most basic level? Recent approaches to causation’s place have been influenced by trends in metaphysics that look to science, rather than intuition, as a guide.

‘Neo-Russellian’ approaches to causation hold that causation is a higher-level phenomenon, whose features are not to be understood by looking primarily to fundamental physics or metaphysics. These accounts draw their inspiration from Bertrand Russell’s critique of causation (1912–3). Russell argued that causal relations don’t appear in how fundamental physical theories are formulated. Fundamental physical theories use laws that relate whole systems at different times: they don’t pick out individual events as causes, and others as effects. If fundamental physics is a direct guide to fundamental ontology, causal relations don’t seem to be part of fundamental ontology. Neo-Russellians argue that Russell’s critique still holds, by and large, for the best contenders for fundamental physics today (Field 2003; Price & Corry 2007). Most, however, don’t think we should eliminate causal notions altogether, as Russell in fact argued, and not all think that fundamental physics settles fundamental ontology.

Some Neo-Russellians think we should explain causation’s place by ‘reducing’, ‘analyzing’ or otherwise showing how causal relations depend on relations that are found in fundamental physics—such as laws and probabilities. Such a reduction would explain how causal relations fit into the picture of the world presented by fundamental physics. According to probabilistic accounts of causation, for example, causal relations derive from objective probabilities. An event a is a cause of a distinct event b, just in case the probability of b is higher given a than otherwise (and other conditions are met). These conditional probabilities may in turn be derived from probabilistic laws or other posits of fundamental physics. For further discussion of probabilistic accounts, see ‘Causation’. 
According to counterfactual accounts, causal relations derive from relations of counterfactual dependence: relations that concern what would be the case, were some part of the world to be different. Some accounts also attempt to analyse counterfactuals in non-causal terms, and so ultimately reduce causal relations to non-causal relations: call these reductive counterfactual accounts. According to David Lewis’s account (1973, 1979), for example, event \( a \) is a cause of a distinct event \( b \) just in case \( b \) counterfactually depends on \( a \), or there is a chain of counterfactual dependencies relating them (for example, \( a \) depends on \( c \), and \( c \) depends on \( b \)). \( b \) counterfactually depends on \( a \), moreover, just in case, if \( a \) were to occur, \( b \) would occur, and if \( a \) were not to occur, \( b \) would not occur. These counterfactual relations are analysed in non-causal terms using physical laws, and the distribution of events in space and time. For more on Lewis’ account, see ‘Causation’, and many of the readings recommended below, particularly Collins, Hall, and Paul (2004).

An alternative Neo-Russellian approach is to accept that causation is primarily a higher-level phenomenon, whose natural home isn’t fundamental physics, but to deny that causal relations need to be reduced to non-causal relations. Prominent examples of this approach are ‘interventionist’ accounts. These relate causation to counterfactuals, but don’t attempt to analyse counterfactuals in non-causal terms. According to accounts defended by James Woodward (2003) and Christopher Hitchcock (2001), a variable \( A \) is causally relevant to another variable \( B \), just in case the probability of \( B \) would change were \( A \) to be intervened on by a suitable (causal) process. What makes something a suitable process is specified in causal terms, so causal relations aren’t reduced to non-causal relations. Instead, different causal and counterfactual relations are related to each other. These accounts build on the ‘structural equations’ framework developed by Judea Pearl (2000) and Peter Spirtes, Clark Glymour, and Richard Scheines (1993). While this framework needn’t be interpreted in counterfactual terms, this has been their most influential interpretation in philosophy.

The primary goal of those defending interventionist accounts is not to determine whether causation is part of fundamental ontology, or how it relates to fundamental physics. Instead, the aim is to provide a framework for deriving causal relations from probabilistic
data and causal assumptions, distinguishing various types of causal relations, and relating causal relations to one another. The framework is also used to examine in what ways causal relations can be sensitive to background conditions (Woodward 2006). For all these purposes, there is no need to derive causal relations from non-causal relations. For further discussion of interventionist accounts, see ‘Causation’.

Another kind of Neo-Russellian approach also puts to one side certain metaphysical questions. According to ‘agency theories’ of causation (Price and Weslake 2009; Ismael 2012), causal relations only appear from the perspective of the deliberating agent who goes around in the world deciding on one thing in order to achieve another. They don’t appear when we take the point of view of fundamental physics. Some agency theorists (Price, and to some extent Ismael) also eschew the project of finding a place for causation in fundamental ontology. They argue, roughly, that the world doesn’t have a fundamental ontology in a metaphysical sense, because we can’t sensibly ask after the nature of the world, independent of any perspective.

Finally, some philosophers reject Russell’s claims altogether. Some argue that causation is a primitive, irreducible relation, part of fundamental ontology, and closely tied to laws (Tooley 1988). Others argue that the world is fundamentally composed of causal dispositions or powers: a kind of Aristotelian view—see Greco and Groff (2013) and ‘Causation’. According to these views, Neo-Russellians misrepresent the content of fundamental physics when they leave out its causal content. Others tie causation much more closely to fundamental physics than standard Neo-Russellians. Philip Dowe (2000), for example, identifies causal relations with particular types of processes that conserve physical quantities.

2. The Temporal Asymmetry of Causation

A characteristic feature of causation seems to be its temporal asymmetry: causes come prior in time to their effects. While many accept the possibility of backwards or simultaneous causation at other possible worlds (such as in time travel scenarios) (Loewer, Lewis), most also accept that causes always precede their effects at our world.
Philosophers have taken different approaches to the temporal asymmetry of causation, depending on their approach to causation’s fundamentality. If causation is fundamental, it may simply be a primitive, not-further-to-be-explained feature of causation that it is directed towards the future. The direction of causation may even define the forwards direction of time. But if causation is not fundamental, there are reasons to think its temporal asymmetry should be explained. As Russell (1912–3) argued, candidate laws of fundamental physics for the most part temporally symmetric, and not substantially different in character towards the past and future. If causation is temporally asymmetric, it seems the laws alone don’t explain its asymmetry. Neo-Russellians defending reductive counterfactual accounts have therefore typically sought to explain causation’s temporal asymmetry.

Lewis (1979) argued that the temporal asymmetry of causation is due to a temporal asymmetry of counterfactuals: if a small local event were different, almost all the past would remain the same, but large parts of the future may well change. Since causal dependence requires a chain of counterfactual dependence, backwards causation is (mostly) excluded. Lewis explained the asymmetry of counterfactual dependence in terms of an asymmetry of ‘overdetermination’: large changes to the present are required to imply a substantially different past (given the laws), but small changes to the present are enough to imply a substantially different future. However, given the time-symmetry of physical laws, it turns out this asymmetry of overdetermination doesn’t hold (Elga 2001). Small changes to the present may imply large changes to the past as well as to the future.

Recently reductionists have improved on Lewis’ approach by drawing on results from statistical mechanics. In Boltzmannian statistical mechanics, one explains temporally asymmetric behaviour (such as why heat disperses in a room) by appealing to temporally symmetric laws and suitable initial conditions—see ‘Time, Direction of’. Albert (2000) and Loewer (2007) argue that the temporal asymmetry of causation can be similarly derived from an initial low-entropy constraint, a probability distribution and time-symmetric laws. Others argue that instead of a low-entropy constraint, one needs to assume initial micro-randomness: in the initial state, positions and velocities of particles aren’t correlated with one another (Frisch 2014, ch. 5).
By contrast, agency theorists attempt to explain causation’s temporal asymmetry by appeal to asymmetric features of agents—such as the fact that we deliberate before we decide, or the fact that we ‘directly’ control future actions (Price and Weslake 2009; Ismael 2012). A further question is whether we need to also explain these temporally asymmetric features of agents in physical terms in order to fully account for causation’s temporal asymmetry. For a somewhat different approach to causation’s temporal asymmetry that involves both decision-making and physics, see Horwich (1987).

Other Neo-Russellians have treated causation’s temporal asymmetry as a contingent fact to be discovered empirically, but not further explained. Interventionists standardly take this approach. Their contribution to the debate has often been to precisify what the causal asymmetry amounts to: that certain patterns of counterfactual (and therefore probabilistic) dependence are seen towards the future and not the past.

3. Counterfactuals, Contrasts and Context
In addition to their fit with science, accounts of causation have also been tested based on whether they agree with our intuitive judgements about core cases. Concerns include i) not counting the wrong correlations as causal, ii) recovering all the causal relations there are, and iii) appropriately distinguishing between causes and background causal conditions. Such concerns are particularly pressing for accounts that derive causal relations in part from counterfactuals or probabilities, although they may also arise as epistemological problems for other accounts concerning how we discover causal relations.

Much of the focus of recent attention has been on ii), and whether counterfactual accounts can deliver ‘actual’ or ‘token-level’ causation: relations between particular causal relata, rather than relations between general types of events. Consider a case. Billy and Suzy are each poised to throw a rock at a bottle. If either throws, the bottle will break. Suzy throws first, and her rock粉碎s the bottle. But, if she hadn’t thrown her rock, Billy would have, and his rock would have smashed the bottle. In this case of ‘early pre-emption’, it seems there is causation (Suzy’s throw causes the bottle to break) but no counterfactual dependence—if Suzy hadn’t thrown her rock, the bottle still would
have broken, due to Billy’s throw. Lewis’ account gets the right result here: there is causation because there is a *chain* of counterfactual dependencies relating the bottle smashing to Suzy’s throw. Suzy’s rock flying through the air depends on Suzy’s throw. The bottle’s smashing depends on Suzy’s rock flying through the air, since, if Suzy’s rock is flying, Billy’s rock would not have been thrown.

But there are cases of ‘late pre-emption’ where Lewis’ response won’t work. Say Suzy and Billy throw their rocks *at the same time*. Suzy’s rock reaches the bottle first, causing it to smash. At any time, if Suzy’s rock hadn’t been there, Billy’s would have, and would have caused the bottle to smash. So there is no chain of counterfactual dependencies relating Suzy’s throw to the bottle smashing. For further discussion of pre-emption cases, see Collins, Paul and Hall (2004, Ch. 1), and Paul and Hall (2013, Ch. 3). Concerns about late pre-emption cases ultimately prompted Lewis (2000) to take up a different ‘influence’ based account of causation.

Interventionists deal differently with cases of pre-emption. One standard move is to hold fixed the ‘off-path’ variables (those not causally between the putative cause and effect) at their actual values. In the early pre-emption case, because Billy doesn’t throw in the actual world, we hold fixed the fact that he doesn’t when evaluating counterfactuals concerning Suzy’s throw. The bottle smashing then depends on Suzy’s throwing. The same strategy, however, doesn’t work in cases of late pre-emption. In the above case, Billy *does* throw in the actual world. If we hold this value fixed, the bottle smashing doesn’t depend on Suzy’s throw. For discussion, see Woodward (2003, Ch. 5). For this reason, interventionists have offered more sophisticated responses. We’ll consider one below. For alternative treatments, see Hitchcock (2001) and Halpern and Pearl (2005).

One way of dealing with cases of late pre-emption is to appeal to a distinction between default and deviant values. Instead of holding fixed off-path variables at their actual values, we hold them fixed at their ‘default’ values—the values that we would ordinarily expect them to take. Given that we don’t expect competing rock throwers to be around, we hold fixed that Billy doesn’t throw. In this case, the bottle smashing does depend on Suzy’s throw. For discussions of this approach see Menzies (2007), and Paul and Hall (2003, Sections 2.2.3, 3.3.3–4).
A default/deviant distinction may play other roles as well, such as distinguishing between causes and background causal conditions (iii). Say the striking of the match causally depends on both its being struck, and the presence of oxygen. The striking might be a *cause*, while the oxygen is a mere *causal condition*, because the striking of the match is deviant behaviour, while the presence of oxygen is default behaviour. A default/deviant distinction can also be used to determine the appropriate contrast sets in ‘*contrastive*’ theories of causation (Schaffer 2005). According to these accounts, it is not properly the case *c* causes *e*, but that *c* rather than *c*\(^*\) causes *e* rather than *e*\(^*\).

Appealing to a distinction between default and deviant values, however, raises further questions. Firstly, what precisely is the distinction? Default behaviour might be what is *statistically* normal for the system (competing rock throwers aren’t typically around), what is in accord with *social norms* (one shouldn’t throw rocks), or what is in accord with normal *functioning* (the rocks are part of a decorative border that should stay put). Each of these alternatives will lead to different results.

Secondly, what determines the appropriate default/deviant distinction? Relatedly, what determines the values the variables take, or the appropriate contrast set? One option is that these are partly determined by *context* (Menzies 2000, 2007; Woodward 2003, Ch. 5; Schaffer 2005). Perhaps in contexts relevant to blame and punishment, for example, the default values derive from social norms, while, in explanatory contexts, the default values are statistical.

Does appealing to context-sensitivity change what we make of the *objectivity* of causation? If context merely helps distinguish between causes and causal conditions, perhaps it is purely a ‘pragmatic’ affair which of the causes we cite. The ‘semantics’ of causation remain fully objective. But if context is required to deliver causal relations in the first place (by distinguishing causation from counterfactual dependence), it seems causal relations themselves are context-sensitive, and less objective than relations of counterfactual dependence. One might then seek to establish what the appropriate values are on context-*insensitive* grounds in order to maintain the objectivity of causation.
Altogether, even very basic features of causation relating to its objectivity remain under dispute—including its context-sensitivity, temporal asymmetry, and place in science and metaphysics.

**References and Further Readings**


(Ch. 6 explains the temporal asymmetry of causation using statistical-mechanics.)


(Recent collection on causation inspired by Peter Menzies’ work, relevant to many issues discussed above.)


(Useful overview, covering a broad range of recent work.)


(Collection on Lewis-style counterfactual accounts of causation, with a focus on problem cases. Ch.1 provides a useful summary.)


(Defends a process theory of causation.)


(Influential critique of Lewis’ explanation of causation’s temporal asymmetry.)


(Accessible overview on causation’s relation to fundamental physics.)


(Argues that causal reasoning is essential, even in physics.)

(Recent collection on powers-based approaches to causation.)


(Defends an interventionist approach to actual causation. Somewhat technical.)


(Defends an interventionist account of causation.)


(Chs. 6–8 discuss causation’s temporal asymmetry, as well as backwards causation.)


(Explains causation’s temporal asymmetry by appeal to agency.)


(Classic presentations of Lewis’ counterfactual account of causation, as well as his revision to it (2000).)


(Defends a counterfactual account of causation using statistical mechanics.)


(Uses a default/deviant distinction to defend an interventionist account of causation.)


(Overview of recent work, focusing on problems for counterfactual accounts.)

(Technical presentation of an interventionist account of causation. Less useful than Woodward (2003) as a starting point.)


(Accessible collection on Neo-Russellian approaches to causation.)


(Defends an agent-based account of causation and its temporal asymmetry.)


(Classic argument against causation featuring in fundamental physics.)


(Argues that causation is contrastive, and explores context sensitivity.)


(Technical presentation of the structural equations framework.)


(Argues that causation is irreducible.)


(Classic presentation of an interventionist account of causation—an accessible starting point.)


(Considers how causal relations are sensitive to background conditions. Relevant to cases discussed in Section 3 above.)