⋆

How can causal explanations explain?

⋆

Jon Williamson

To appear in *Erkenntnis*
Draft of July 16, 2013

Abstract

The mechanistic and causal accounts of explanation are often conflated to yield a 'causal-mechanical' account. This paper prizes them apart and asks: if the mechanistic account is correct, how can causal explanations be explanatory? The answer to this question varies according to how causality itself is understood. It is argued that difference-making, mechanistic, dualist and inferentialist accounts of causality all struggle to yield explanatory causal explanations, but that an epistemic account of causality is more promising in this regard.

According to mechanistic philosophy of science, scientists explain a phenomenon by pointing to the mechanism responsible for it (see, e.g., Machamer et al., 2000). This seems to offer a good account of many scientific explanations, which do indeed tend to depend heavily on describing underlying mechanisms. On the other hand, Lewis (1986a) and others have argued in favour of a *causal* account of explanation: according to this account, we explain an event by pointing to the chains of causes that led up to it.

This paper asks how, if the mechanistic account of explanation is essentially correct, causal explanations can be genuinely explanatory. If the mechanistic account is correct, causal claims only explain to the extent that they shed light on mechanisms. Here mechanisms are understood broadly to include not only the fixed hierarchical structures of components interacting in such a way as to regularly produce some phenomenon (often called *complex-systems mechanisms*) but also low-level physical processes (*Salmon-Dowe mechanisms*); see §1. The extent to which causal claims shed light on underlying mechanisms depends on the precise account of causality invoked (§2). This paper evaluates difference-making accounts of causality (§3), mechanistic accounts (§4), dualist accounts (§5), certain inferentialist accounts (§6), and the epistemic theory of causality (§7), to see whether any such account can successfully render causal claims explanatory. It is argued that the epistemic theory is most promising in this regard.

§1
The mechanistic account of explanation

The mechanistic account of explanation is the cornerstone of the recent interest in mechanisms in the philosophy of science. Thus Machamer et al. (2000) begin their paper with:

> In many fields of science what is taken to be a satisfactory explanation requires providing a description of a mechanism. So it is not surprising that much of the practice of science can be understood in terms of the discovery and description of mechanisms (Machamer et al., 2000, pp. 1–2).

Mechanistic accounts of explanation have also been put forward by Salmon (1984, 1998); Glennan (2002); Bechtel and Abrahamsen (2005); Craver (2007) and others.

Note that different authors have different things in mind when they talk about mechanisms. One school of thought has it that mechanisms need to be understood as *physical processes*, i.e., spatiotemporally contiguous processes in which a mark or a conserved quantity is propagated between interactions (Reichenbach, 1956; Salmon, 1984, 1998; Dowe, 2000). An example of this sort of mechanism is a signal from a remote control to open a garage door: pressing the button constitutes an interaction which leads to the transmission of a signal that is propagated in such a way that it can interact with a receiver at the garage. An alternative to the physical-process view is the idea of *complex-systems mechanisms* (CSMs). These consist of entities and activities organised in such a way that they are responsible for some phenomenon (see, e.g., Machamer et al., 2000; Illari and Williamson, 2012). An example is the remote control mechanism itself, responsible for sending the signal that opens the garage door: this is a more-or-less stable arrangement of parts that can engage in characteristic activities that lead to the transmission of the signal.

These views need not be construed as alternatives. One can also take a broad view of mechanisms, according to which mechanisms involve physical processes or complex-systems mechanisms or some combination of the two. An explanation of the garage door opening might then describe or point to: (i) the CSM for producing the signal; (ii) the physical signal itself; and (iii) the CSM for receiving the signal and opening the door.

Note that two types of explanation are possible: *single-case*, i.e., a *particular* garage door opening is explained by (i–iii) together with the particular fact that the remote control was triggered in the appropriate way; or *generic*, i.e., garage door openings *in general* are explained by (i–iii). Most of the following discussion will apply to both single-case and generic explanation.

A second distinction is also useful. An *explanation in practice* is a communication that aims to increase the understanding of an interlocutor by describing how an explanandum (a single-case event or a generic phenomenon) is produced by underlying mechanisms that the interlocutor understands or accepts better than the explanandum itself. On the other hand, an *ideal explanatory text* is an imaginary text that would recursively describe all the underlying mechanisms: i.e., that includes descriptions of the mechanisms that are responsible for the explanandum, other mechanisms that are responsible for the appropriate functioning of those mechanisms, and so on. The concept of an ideal explanatory text faces the *bottoming-out problem*: some account needs to be given as to whether there is a lowest level of

mechanism, and, if so, what explains the proper functioning of those mechanisms. We will chiefly be concerned with explanation in practice in this paper.

The mechanistic account of explanation is the object of ongoing research and development—as yet there is no consensus as to how the bottoming-out problem should be solved, for example. However, the account as it stands already seems to be true to much of our explanatory practice, including scientific practice. The teaching of science is the most striking example of mechanistic explanation in practice: witness the plethora of textual and pictorial descriptions of mechanisms in many scientific textbooks. 'The biology of cancer' (Weinberg, 2007), for example, contains descriptions of mechanisms on almost all of its 800 pages. (Admittedly textbooks at the pure end of the mathematical sciences make less use of mechanisms, but that is because they do not normally seek physical explanations of mathematical phenomena.) But it is not just pedagogical science that invokes mechanisms: much of scientific research, to the extent that it seeks to explain rather than describe, predict or control phenomena, proceeds by invoking mechanistic explanations. For example, the envirogenomarkers project (http://www.envirogenomarkers.net) seeks to explain disease in terms of environmental exposure by appealing to (i) the physical processes that bring the environmental factors to the human body; (ii) the eventual failure of the body's CSMs for preventing damage; (iii) resulting processes that lead to disease (see, e.g., Chadeau-Hyam et al., 2011; Russo and Williamson, 2012). This is a process–CSM–process explanation rather than the CSM–process–CSM explanation that we encountered when explaining the opening of the garage door, but it is clearly in the same mechanistic mold.

It is no wonder that so many of our explanations are mechanistic, for the explanatory virtues of mechanistic explanations are fairly obvious. Most fundamentally, in a mechanistic explanation it is clear that the explanans can *increase one's understanding* of the explanandum, especially if one already understands or accepts the physical processes involved as well as the entities, activities and organisation of the CSMs invoked in the explanation. Mechanisms are *real*, so there is no question about the ability of the mechanisms to explain, as there is with some counterfactual explanations whose truthmakers are more speculative. A mechanism is also *local* to a particular part of the world, so the phenomenon is explained in terms of the make-up of the part of reality in which it occurs; on the other hand, many regularity or law-based explanations explain things in terms of the pattern of occurrence across this world and perhaps other possible worlds (Illari and Williamson, 2011).

In sum, the mechanistic account of explanation is a powerful and compelling approach, and it appears to be essentially correct at least for a large swath of explanatory practice.

§2
The causal account of explanation

In contrast to the mechanistic account of explanation, which proceeds by explaining some phenomenon by pointing to the mechanisms responsible for it, the causal account of explanation explains an event by pointing to a chain of events—or network of events—that led up to it (e.g., Lewis, 1986a). Thus in response to a question of the form: *Why did / does E occur?* the causal account says: *Because C occurred / occurs and C is a cause of E.* Or, more generally: *Because $C_1,\ldots,C_n$ occurred / occur and $C_1,\ldots,C_n$ are causes of E via network n.*

The question arises as to how, if the mechanistic account of explanation is essentially correct, a causal explanation can be explanatory. This is the key question of this paper. The answer to the question will clearly depend very much on what 'causes' means: some accounts of causality may render causal explanations more explanatory than others. In what follows we shall investigate various accounts of causality in order to assess the extent to which they yield causal explanations that are genuinely explanatory. While there isn't the space here to investigate any particular account in any detail, a survey of a variety of accounts will give us the opportunity to draw some comparisons.

Note that the mechanistic and causal accounts of explanation have sometimes been conflated and referred to as the 'causal-mechanical' account of explanation (see, e.g., Salmon, 1998, §4.3; Woodward, 2003, Chapter 8). This conflation is rather natural for those who regard causality and mechanisms as inter-reducible—e.g., Salmon adopted a mechanistic account of causality; Woodward endorses a causal account of mechanisms. However, the two accounts of explanation must be prized apart if we are to address the key question posed above. It is clear that Lewis-style causal explanation is characterised in a very different way to process/CSM-style mechanistic explanation as outlined in §1. If we are to say how the former sort of explanation can satisfy the demands of the latter sort of explanation, we must take these differing characterisations as our starting point.

Most extant philosophical theories of causality can be classified according to whether they are *difference-making*, *mechanistic*, *dualist* or *inferentialist* accounts, and we shall now examine these kinds of account in turn.

§3
Difference-making accounts of causality

The core idea behind difference-making accounts of causality is that causal relationships are to be analysed in terms of patterns of difference making: C is a cause of E iff C makes the appropriate sort of difference to E. As to what the appropriate sort of difference is, this varies somewhat from account to account, but all accounts require that there should be some chain of events $A_0 = C \longrightarrow A_1 \longrightarrow A_2 \longrightarrow \cdots \longrightarrow A_{n-1} \longrightarrow A_n = E$ from C to E such that each member A_i of the chain changes the probability of the next member A_{i+1}.[1] Standard probabilistic accounts of causality tend to require that A_i changes the probability of A_{i+1} conditional on some state of A_{i+1}'s other causes (Williamson, 2009). Counterfactual accounts require that there be a chain of counterfactual dependence between C and E, such that for each link on the chain, if A_i had not occurred, then A_{i+1} would have had a much lower chance of occurring (Lewis, 1973). Finally, agency accounts (also called 'interventionist' accounts) require that intervening to change A_i should change the probability of A_{i+1}, for some fixed state of the A_{i+1}'s other causes (Price, 1991; Menzies and Price, 1993; Woodward, 2003).

Coupling a difference-making account of causality with a causal account of explanation gives, to the question *Why does / did E happen?*, the answer *Because C happens / happened and there is a chain of events from C to E, such that each event on the chain makes / made a difference to the probability of the next event.*

[1] There may be more than one such chain from A_0 to E—i.e., a *network* of events—but there should be at least one such chain.

Recall that the key question of this paper is this: if the mechanistic account of explanation is essentially correct, how can causal explanations be explanatory? When asked, *Why did the garage door open?*, the mechanistic account of explanation would point to the underlying CSM-process-CSM mechanism structure. If that structure were better understood or accepted than the phenomenon to be explained, it might stop there, otherwise it would seek to recursively explain aspects of the CSMs or process involved. On the other hand, the causal account of explanation currently under consideration would answer: *Because the button was pressed and there is a chain of difference making from that press to the garage door opening.* This sort of explanation isolates a relevant factor—the pressing of the button—but says nothing yet about the connection between this factor and the explanandum, other than that this connection involves a chain of difference making, i.e., of the appropriate sort of probability changing. One might ask: *What* is the chain of difference making? This would be answered by providing more detail: e.g., the pressing of the button made a difference to the signal being sent from the remote control, which made a difference to the signal being received by the garage door, which made a difference to the garage door opening. But one might also ask, for any particular link from A_i to A_{i+1} on this chain, *Why* is there this difference making? The causal account of explanation would of course offer an explanation of this form: Because an event B occurred and there is a chain of difference making from B to the event that A_i makes a difference to A_{i+1}. Again, this isolates a further relevant factor.

From the point of view of the mechanistic account of explanation, this process of providing more and more detail to the network of difference making is less than illuminating, for the simple reason that at best it highlights milestones in the mechanisms responsible for the phenomenon, rather than the organisation and arrangement of parts and activities and the corresponding transfer of energy and other conserved quantities. The mechanist would say that it is organisation and energy transfer that often does the explaining, not particular points in the history of the mechanisms involved. An explanation that appeals to the thickness of a cell membrane or the duration of an action potential cannot readily be couched in terms of causal chains of events. To put the problem another way: to say that there is difference making is just to say that things tend to happen that way; to describe the underlying mechanisms is to show why they happen that way.

In sum, from the point of view of the mechanistic account of explanation, difference-making accounts of causality are problematic because they yield impoverished explanations that consist of a discrete network of events rather than a description of the underlying mechanisms in all their structural richness.

Of course, many argue that difference-making accounts of causality are problematic anyway. First, because there can be causation without difference making. If all roads lead to Rome then although taking the high road will cause you to get to Rome, it won't make a difference to your getting to Rome.[2] Second, this sort of account does not seem to accord well with the epistemology of causality. If a difference-making account were correct then establishing the appropriate difference-making relationships should suffice to establish a causal claim. However, as a matter of fact evidence of a mechanistic connection between the putative cause and putative effect is often required over and above excellent evidence of difference

[2]One might think that such cases of overdetermination can be dissolved by appealing to *chains* of difference-making involving intermediate events, but there are analogous particle-decay examples where the prospect of finding suitable intermediate events is rather dim (see, e.g., Williamson, 2009, §9).

making (see, e.g., Russo and Williamson, 2007). This is explicit, for instance, in the conditions for causal discovery provided by Bradford Hill (1965).

Given these problems facing difference-making accounts of causality, it is natural to turn to a mechanistic account of causality in order to attempt to say how causal explanations can be explanatory.

§4
Mechanistic accounts of causality

The core idea behind mechanistic accounts of causality is that causal relationships are to be analysed in terms of mechanistic connection; C is a cause of E iff there is the appropriate sort of mechanism (or chain of mechanisms) linking C and E. As to what the appropriate sort of mechanism is varies from account to account, but varies in line with the discussion of the mechanistic account of explanation in §1. One approach is to require that a mechanism be a physical process that carries a mark or conserved quantity (Reichenbach, 1956; Salmon, 1984, 1998; Dowe, 2000). A second approach is to invoke CSMs (see, e.g., Glennan, 1996). A third alternative is to be more liberal and consider both sorts of mechanism as admissible.

Coupling a mechanistic account of causality with a causal account of explanation gives, to the question *Why does / did E happen?*, the answer *Because C happens / happened and there is a chain of mechanisms of the appropriate sort linking C to E*. With regard to the key question of this paper (how, if the mechanistic account of explanation is right, can causal explanations be explanatory?) the important point to note is that this kind of causal explanation is essentially an *existential* claim: to invoke C as an explanation of E is just to say that *there are* certain mechanisms linking C and E, not what they are, nor what C's role is in these mechanisms. This is unsatisfactory from the perspective of the mechanistic account of explanation, because to say that there is a chain of mechanisms for the garage door opening from the button press is just to say that there is an explanation, not to give the explanation. To be sure, as in the case of difference-making accounts of causality, providing more detail regarding the causal history of the garage door opening can help here, by providing more information about the linking mechanisms. But only partial information: again, at best this detail provides a set of milestones for the progress of the pertinent mechanisms, not the mechanisms themselves. Causal explanations remain impoverished.[3]

Of course many argue that mechanistic accounts of causality are problematic anyway. First, there can be causation without linking mechanisms. Failing to press the button can cause the door to fail to open, without there being any physical connection between these absences.[4] Second, there is the problem of according with the epistemology of causality. If the mechanistic account of causality were correct then it should suffice to establish the existence of a chain of linking mechanisms in order to establish a causal relationship. But typically more is required: even when the mechanism is well known, the causal claim is not established until there is

[3] Darden (2013) also argues that causal explanations are impoverished in comparison to mechanistic explanations, by analysing cystic fibrosis as a case study.

[4] The standard response of the mechanist to this problem of causation between absences is to say that the causal claim is made true by the fact that the expected mechanism didn't operate. This move yields pluralism: a mechanistic account of causation between presences and a counterfactual account of causation between absences (had the button been pressed then the door would have opened). This sort of response is criticised in Williamson (2011) and in §5 below.

also good evidence that the putative cause makes a difference to the putative effect (Russo and Williamson, 2007).

For these reasons, an appeal to a mechanistic account of causality offers little more than a blind alley for the proponent of a mechanistic account of explanation who seeks to understand how causal explanations can be explanatory.

§5 Dualist accounts of causality

We saw that causality without difference making poses one problem for difference-making accounts of causality, and causality without mechanisms presents a problem for mechanistic accounts of causality. A natural response to these problems is to advocate a form of pluralism according to which causal claims are ambiguous: some causal claims are really claims about difference making, while others are claims about mechanistic production (Hall, 2004). Dualist accounts of causality vary according to the particular difference-making account and the particular mechanistic account to which they appeal.

How can causal claims be explanatory on this dualist account, given the mechanistic account of explanation? Saying 'Because C is a cause of E' in response to the question 'Why did E occur?' is ambiguous. It might be saying something about a chain of difference making or it might be saying something about a chain of mechanisms (or, in some cases, about both). In the first situation we inherit the problem that difference-making accounts of causality yield explanations that are impoverished in certain respects. In the second situation we inherit the problem that mechanistic accounts of causality yield explanations that are impoverished in certain other respects. But worse, we now don't know—in the absence of further disambiguating information—which situation we are in. So dualism multiplies, rather than resolves, the explanatory problems of the difference-making and mechanistic accounts of causality.

Dualism arguably also fails to solve the other problems that we saw beset the difference-making and mechanistic accounts of causality. First, if cases of overdetermination provide counterexamples to difference-making accounts of causality and cases of causation between absences provide counterexamples to mechanistic accounts of causality, then cases of overdetermined causation between absences provide counterexamples to dualism, since these are cases of causation with neither difference making nor mechanistic connection (Longworth, 2006b, §4.1). Second, the epistemological problems remain: dualism cannot account for the need for both evidence of difference making and evidence of mechanistic connection in order to establish a causal claim. Consider any particular causal claim that is interpreted as a claim about difference making but not about mechanistic connection. Here it would seem that, once the difference making were established, there would be no need for evidence of a mechanistic connection. On the other hand, in the case of a causal claim that is interpreted mechanistically but not in terms of difference making, evidence of difference making would be irrelevant if the mechanistic connection is known. So we see, then, that dualism inherits the epistemological problems of each of the two sorts of account of causality to which it appeals.

To add insult to injury, there is a lack of evidence for the ambiguity of causal claims. In the case of probability, it seems clear that certain claims of the form *the probability of E is x* can be used to either talk about relative frequency or about degree of belief. When faced with such a claim, the response is sometimes to

disambiguate, either by analysing the context in order to determine whether the probability is single-case (favouring a degree of belief interpretation) or generic (a frequency interpretation), or by asking a clarifying question such as, 'What do you mean? Are you talking about a proportion or your degree of belief?' This provides evidence in favour of dualism about probability. But similar evidence seems to be lacking in the case of causal claims. While it would be entirely appropriate in normal linguistic usage to seek to disambiguate 'the bank is riddled with vermin', a clarifying question would seem out of place when faced with a claim such as 'smoking is a cause of cancer'.

This dualist position is a form of pluralism in which causation is analysed in terms of two determinate relations, a difference-making relation and a mechanistic relation. Given the problems that beset dualism, it may seem natural to advocate more nebulous versions of pluralism (Williamson, 2006). We shall consider one such version now.

§6
Inferentialist accounts of causality

One view of causality has it that causality is a cluster of concepts, standing in some loose family resemblance relation (Anscombe, 1971; Cartwright, 2004; Longworth, 2006a; Psillos, 2009). An interesting development is provided by Reiss (2011), who marries this Wittgensteinian pluralism with contemporary inferentialism. According to this view, the meaning of a causal claim is constituted by its inferential base (the propositions from which one infers the claim) and its inferential target (the propositions that one infers from the claim). There are two variants of this view, according to whether the inferences are given a descriptive or a normative reading. On the descriptive reading, the inferential base and target are determined by actual practice: they are respectively the set of propositions from which a community infers the claim and the set of propositions which the community infer from the claim. On the normative reading, the inferential base consists of the propositions from which one ought to infer the claim and the inferential target consists of the propositions that one ought to infer from the claim.

Note that, as described above, the descriptive and the normative views are not mutually exclusive. The descriptive view often motivates a weak normative view according to which the actual practice of the particular community in question determines what one ought to infer. One line of argument is this: the meaning of a causal claim is constituted by the actual inference base and target as per community usage; therefore, if you want to use causal claims in accordance with their meaning, you had better accord with the actual usage—i.e., the propositions from which you ought to infer the claim are those in the actual inference base and the propositions that you ought to infer from the claim are those in the actual inference target. Another form of the view has it that the normative attitudes or commitments of the community directly determine the inference base and target. Either way, this weak normative view can be distinguished from a strong normative view which invokes some independent standard that determines what one ought or ought not infer. I will use the term *weak inferentialism* to denote the descriptive or the weak normative view, according to which the inferential base and target are determined by actual usage or commitments of the community, and *strong inferentialism* to denote the strong normative view, according to which something other than communal practice

determines what constitutes an appropriate inferential base and inferential target. (Wittgenstein's own position appears closest to weak inferentialism.)

The key difficulty for strong inferentialism is that the inferentialist component appears to be redundant. This view appeals to some independent standard that determines what one ought or ought not infer, in order to yield some facts of the form *one should infer that C is a cause of E from proposition θ*, and, *one should infer proposition φ from the claim that C is a cause of E*. But then one could simply say this: in a world whose state can be captured by proposition θ, whatever makes it true that *one should infer that C is a cause of E from θ* makes it true that C is a cause of E. In sum, the worry is that talk of inference is just beating about the bush—inference is doing little or no work: the account of causality is essentially being provided by the independent standard of inferential success.

Reiss himself steers towards weak inferentialism:

> there aren't many hard-and-fast rules that philosophers can use to prescribe scientists and ordinary folk what inferences they should and shouldn't make. The best guide to what's doable and what isn't is scientific practice and therefore I won't make highly general claims about what a language user is entitled to. (Reiss, 2011, p. 916.)

While it seems very sensible to look to scientific practice to set a standard for causal inferences, this weak inferentialist view has its own share of problems.

First, it leads to a very radical form of pluralism. It deems the word 'cause' to mean different things in 'inhaling tobacco smoke is a cause of cancer in mice' and 'inhaling tobacco smoke is a cause of cancer in humans'. This is because the inferential base is very different for the two claims: scientific practice deems that randomised controlled trials may occur in the inferential base of the former claim, but not in the inferential base of the latter, for purely ethical reasons. Now under an inferentialist account, if two causal claims differ as to the kinds of propositions that occur in the inferential bases or inferential targets, they are claims about two different kinds of causality. Reiss is explicit about this:

> it is easy to see how this theory of meaning leads to a form of pluralism about causation. If its inferential connections to other propositions constitute the meaning of a causal claim and the kinds of propositions from which a causal claim can be inferred and those that can be inferred from a causal claim differ from claim to claim, the case for pluralism has been made. Very roughly, we can define identity conditions for causal claims as follows. Suppose the term 'cause' is used on two different occasions and it is not known whether it has the same meaning on both occasions. Two such claims would have the form 'X α-causes Y' and 'Z β-causes W'. We can then say that 'α-causes' has the same meaning as 'β-causes' (on these occasions) to the extent that 'X α-causes Y' is inferentially connected to the same kinds of propositions regarding the relation between X and Y as 'Z β-causes W' is inferentially connected to propositions regarding the relation between Z and W. If, to give a fictional example, both 'X α-causes Y' and 'Z β-causes W' have been established by RCTs and both license claims about effective strategies (such as 'promoting X is an effective means to raise the chance of Y' and likewise for Z and Y), then 'α-causes' means the same as 'β-causes' (on these occasions).

> There is no guarantee that the kinds of propositions found in inferential base and target are the same for different instances of 'cause'. Different methods of supporting a causal claim license different kinds of inference. (Reiss, 2011, pp. 923–924.)

But it is simply not plausible that different senses of cause are involved in 'inhaling tobacco smoke is a cause of cancer in mice' and 'inhaling tobacco smoke is a cause of cancer in humans'. Indeed, it appears that the underlying mechanism is more or less the same in each case, and that we have difference making in each case, so even the dualist would not distinguish different senses of cause here. If there is scant evidence for the pluralism generated by dualist accounts of causality, the pluralism generated by inferentialism is in a far worse position.

The second key problem for a weak, Wittgensteinian inferentialist account of causality is to do with fallibility. If the meaning of 'cause' is determined by the usage of the word in our inferential practice, how can it be possible for us to make systematic mistakes when we draw causal inferences? If we all use the word 'cause' in the same way, or at least have the same normative commitments regarding such usage, then the inferentialist cannot say that inferences to or from causal claims that accord with these communal standards are erroneous or invalid. Any change in the inferential base or target would lead to a different notion of cause, not better usage of the same notion of cause. This holds just as much for scientific usage as it does for folk usage (Lakatos, 1978). Today there is some some suggestion that scientific practice of evidence-based medicine (EBM) may be systematically mistaken because the evidence hierarchies that the EBM movement invokes underplay the role of mechanistic evidence (Russo and Williamson, 2011; Clarke et al., 2012). If correct, this would indicate certain scope for improvement with respect to the inferential base for causal claims (equivalently, improvement with respect to the inferences that one is entitled to draw from a given set of evidence). Moreover, some have suggested that when we apply causal claims produced on the basis of evidence involving one population to, say, inferences about public policy with regard to another population, we are routinely misapplying these causal claim (see, e.g., Cartwright and Efstathiou, 2011). If correct, this problem of external validity would indicate scope for improvement with respect to the inferential target of causal claims. But under a weak inferentialist account there is no scope for *improvement* of inferential base or target, for changes in base or target simply change the concept of cause altogether. The only way for the inferentialist to capture the notion of improvement is to appeal to some independent normative standard; then one can say that a change in the concept of cause is an improvement to the extent that the new concept invokes more appropriate kinds of inferential base and target. But then we are back to square one: the truthmakers of these normative claims are doing all the work in characterising our concept of cause.

So while strong inferentialism faces a threat of redundancy, weak inferentialist accounts of causality suffer from being implausibly pluralist and from an inadequate response to the possibility of systematic error. On the other hand, they do yield causal explanations that are explanatory, at least in a trivial sense. It is a fact that our causal claims are used to generate explanations. I.e., explanations are in the inferential target of many causal claims. Now an inferentialist concept of cause is determined by the inferential base and target. (There are many such concepts because the inferential base and target can vary quite substantially.) So any concept of cause that has explanations in its inferential target cannot but have explanations

in its inferential target. Such a concept cannot but be explanatory.

Of course this does little to answer our question of how—if a mechanistic account of explanation is correct—causal explanations can be explanatory. It is not part of the normal usage of causal claims that they generate explanations that are like mechanistic explanations, appealing to organisational or structural features such as cell membranes. Rather, causal explanations tend to invoke a history of discrete events.

Moreover, there is a further problem for causal explanations on a weak inferentialist account of causality. As we saw above, it is apparently impossible, under such a view, for our explanatory practice to be routinely mistaken. Similarly, such a view makes it apparently impossible to improve the way we infer causal explanations from causal claims, because if we change our explanatory practice we change our notion of cause, and a weak account has no independent standard by which to deem one concept of cause an improvement over another. To those who think it plausible that there is scope for improvement in science, where explanation plays a leading role, this is a rather perverse position.

So our quest to understand how causal explanations can be explanatory must continue, and we turn next to the epistemic account of causality. While for the inferentialist the inferential base and target act together to determine the meaning of a causal claim, under the epistemic account of causality the roles of the inferential base and target are strictly separated. As we shall see, the inferential target is used to provide an independent normative success criterion for the way in which causal claims are posited from the inferential base.

§7
The epistemic account of causality

According to the epistemic account (see, e.g., Williamson, 2005, Chapter 9), causality is a feature of the way we represent reality rather than a feature of agent-independent reality itself; it is neither reducible to patterns of difference making nor to physical mechanisms. Our causal beliefs help us with our dealings with the world, since they allow us to predict, to explain and to control reality. We have these causal beliefs because of this inferential utility, not because there is some non-epistemic causal relation that is the object of those beliefs.

Causal beliefs can be understood analogously to Bayesian probabilistic beliefs. Just as, under the Bayesian account, a probabilistic belief is a kind of belief (a degree of belief of the form $P(E) = x$) rather than a belief about a non-epistemic probability, a causal belief is a kind of belief (a directed belief of the form $C \longrightarrow E$) rather than a belief about a non-epistemic causal relationship. Just as, under the Bayesian account, we have probabilistic beliefs in order to draw inferences about the world (to help us bet on events), we have causal beliefs to draw inferences about the world (to predict, explain and control the world).

It is the inferential target that plays a crucial role in each case. On the one hand it delimits the scope of the concept: degrees of belief are that which act as betting quotients; causal beliefs are that which ground certain kinds of prediction, explanation and control (PEC) inferences. On the other hand the inferential target provides a success criterion: degrees of belief are rational to the extent that they lead to successful bets (typically in the sense that they lead to bets that do not leave one prone to avoidable losses); causal beliefs are rational to the extent that they lead to successful PEC-inferences. The inferential target is thus an independent

standard by which we can judge the use of the inferential base. A probabilistic epistemology, i.e., a set of inferences from evidence (inferential base) to degrees of belief, is appropriate to the extent that it yields inferentially successful degrees of belief. Similarly a causal epistemology, i.e., a set of inferences from evidence to causal beliefs, is appropriate to the extent that it yields inferentially successful causal beliefs. This gives us the concept of an *ideal* probabilistic (respectively, causal) epistemology—that which yields an optimally successful set of probabilistic (respectively, causal) beliefs.

The epistemic theory of causality then characterises the causal relation thus: C causes E iff there is an ideal causal epistemology which, when applied to an evidence base consisting of all fundamental matters of fact, yields that causal claim.[5]

Note that the epistemic account of causality does not succumb to the two key problems that beset the weak inferentialist accounts. First, the epistemic account does not suffer from the radical pluralism that faces weak inferentialism. The word 'cause' means the same thing in 'inhaling tobacco smoke is a cause of cancer in mice' and 'inhaling tobacco smoke is a cause of cancer in humans'. This is because under the epistemic account the causal relation is characterised purely in terms of the inferential target—PEC-inferences—and not the inferential base, so causal claims with different sorts of inferential base are not claims about different sorts of cause. Neither is there any pluralism induced by the inferential target. Although as a matter of fact certain causal claims may be used chiefly for explaining, while others are used for prediction or control, this does not imply that these are claims about different notions of cause. This is because any causal claim *grounds* prediction inferences, explanation inferences *and* control inferences, even if in practice we lack the inclination or further information required to make—or make use of—some of these inferences.

Second, the epistemic account does not suffer from the problem of capturing the notion of improvement that we saw was a problem for a weak inferentialist account of causality. This is because the inferential target provides the required independent success criterion. Thus a current causal epistemology may be flawed because it fails to attach the appropriate weight to a certain sort of evidence. Another causal epistemology, which weighs evidence differently, may improve on it, as long as it yields sets of causal claims that generate more successful PEC-inferences. Changing the epistemology does not change the concept of cause, because the concept of cause is tied solely to the inferential target, not the inferential base.

On the other hand, changing the inferential target does change the concept of cause. But that does not mean that the epistemic theory cannot account for improvement. While it is not possible, under the epistemic account, to be systematically mistaken about the uses to which one puts causal relationships, it is possible for a new concept of cause to improve on an old concept if its PEC-inferences are more successful than those of the old concept. There is a sense, then, in which we might be making PEC-inferences that are too bold, if the problem of external validity is as bad as some fear: this would be the case if, were we to draw more cautious PEC-inferences, they would be more successful than at present. In such a situation it would not be that we mistakenly use the current notion of cause, but rather that there is a better notion of cause to which we should strive.

[5] This characterisation leaves room for a certain amount of subjectivity as to what causes what, where two ideal causal epistemologies disagree. As with Bayesian accounts of probability, there is scope for different epistemic accounts of causality to yield differing degrees of subjectivity. Accounts towards the objective end of the spectrum are defended in Williamson (2005) and Williamson (2010).

One might try to adopt an alternative strategy here and say that causal relationships are characterised, not by the current PEC-inferences, but by some ideal PEC-inference relation. Then it would be possible to be systematically mistaken about the uses to which one puts causal relationships, insofar as current PEC-inferences fail to live up to the ideal. The question as to whether this is an appropriate strategy depends on whether one is trying to characterise our *concept* of cause, or some metaphysically independent causal relation. The advantage of appealing to current—rather than ideal—PEC-inferences in characterising the concept of cause, is that it offers a natural way of capturing the phenomenon of conceptual change. If characterised in terms of ideal PEC-inferences, the concept of cause would be immutable. Yet the concept of cause does seem to change: the Aristotelian notion of cause had a far wider range of uses than the current notion and admitted a pluralist reading; furthermore, the concept arguably changed in the 20th century when it became associated more with probabilistic predictions as universal determinism fell out of favour and as claims in medicine (such as *smoking causes lung cancer*), that apparently fail to admit deterministic predictions, became more pervasive. If we want to capture conceptual change we need to characterise the causal relation of the moment, and the appeal to the PEC-inferences of the moment is a natural way to do this.

In sum, then, new concepts of cause are *better* concepts of cause to the extent that the new PEC-inferences are more successful. They are concepts of *cause* to the extent that the new inferences are PEC-inferences at all, rather than, say, just prediction-inferences.

This brings us to the question of success. When are PEC-inferences to be deemed successful? Prediction and control inferences seem least problematic. Prediction inferences are successful to the extent that the predictions are borne out. Control inferences are successful to the extent that one can actually achieve one's goals by intervening in the recommended way. But what is it for a causal explanation to be successful? The mechanistic account of explanation can be taken as an independent standard here (§1): a causal explanation is successful to the extent that it elucidates the mechanism responsible for the effect to be explained. On the epistemic account, causal explanations can be judged to be explanatory insofar as they latch onto mechanistic explanations.[6]

We saw above that difference-making, mechanistic and dualist accounts of causality face certain problems in this respect. Because they deem the causal relation to be 'out there', independent of epistemic considerations, causes and effects are constrained to be certain sorts of entities, normally spatiotemporally disjoint events. Correspondingly, causal explanations are constrained in the sense that they chart event histories, rather than mechanism structure and organisation. The epistemic account is not subject to the same constraints. The epistemic account neither analyses causality in terms of chains of probabilistic dependence, nor in terms of chains of mechanistic connections. So causes and effects need not be spatiotemporally disjoint events. They can be variables, absences, properties, facts, entities, activities, organisation, or indeed other causal relations (Williamson, 2005, Chap-

[6]For the mechanistic account of explanation to offer an independent standard of explanatory success, one would not want to go on to analyse mechanisms in terms of causality, for fear of circularity. This is not to say, however, that mechanisms should not invoke causal relationships. Rather, mechanisms should not bottom-out by appealing to epistemic causality. Instead, whatever is taken to explain the workings of the lowest-level mechanisms should be other than causal beliefs—basic activities, dispositional properties or fundamental laws will do, for example.

ter 10)—whatever is needed to ground successful PEC-inferences. Thus under the epistemic account it is in principle possible for causal explanations to offer richer descriptions of the salient aspects of mechanistic explanations.

We saw that difference-making, mechanistic and dualist accounts of causality suffer from two further problems: counterexamples, and a failure to account for the need for both mechanistic and difference-making evidence for causal claims. These problems don't beset the epistemic account of causality. To say that there is a counterexample is to say that the epistemic account deems $C \longrightarrow E$ to be (respectively, not to be) a causal relationship when in fact it isn't (is). But this is to say that such a mistake would be made by every ideal causal epistemology when applied to evidence consisting of all matters of fact. Short of invoking malevolent demons, it is hard to see how this would be possible. Turning now to the evidence problem, where there is a need for evidence of both mechanistic connection and difference making in order to establish a causal claim, an ideal causal epistemology would only yield a causal claim in the presence of both sorts of evidence. So the epistemic account is well-suited to this sort of evidential pluralism.

We see, then, that the epistemic theory overcomes some of the difficulties facing other accounts of causality, and that there is scope for the epistemic theory to account for how, when causal explanations are genuinely explanatory, this can be so.

§8
Summary and further questions

In this paper we have set the mechanistic account of explanation against the causal account of explanation and seen that, if the former is essentially correct, several extant accounts of causality struggle to yield causal explanations that are genuinely explanatory. While this survey has suggested that the epistemic account seems most promising in this regard, the discussion has necessarily been conducted at quite a general level and it would be interesting to look in detail at case studies to shed further light on the question.

One might ask, why give causal explanations at all if mechanistic explanation is the standard to which to strive? Simplicity might be the answer. A flick through any biology textbook is enough to demonstrate the complexity of many mechanistic explanations. Explanation in practice, though, often requires a more cursory sketch of the features of the mechanism that are salient both in terms of the difference they make and in terms of the prior knowledge of the target audience. This is where causal explanation comes into its own. The causal relation is a simple binary relation; causal relationships are simple to communicate, and since we all have causal beliefs, relatively simple to assess and integrate into our epistemic states if necessary.

But much is expected of this humble binary relation. It is overloaded with a multiplicity of uses: prediction, explanation and control. An ideal causal epistemology must thus make some compromises between simplicity, explanatory success, predictive success and interventional success. And it does this on the basis of evidence of difference making, mechanisms, temporal cues, and so on. One advantage of inferentialist accounts and the epistemic account is that they seek to remain true to this rich tapestry of connections between causal claims and their evidence and uses.

While the epistemic account has been juxtaposed here against inferentialist accounts of causality, and compared to Bayesian accounts of probability, it can also be thought of as analogous to the best-system view of laws. There is a sense in which the causal relation is a compromise of simplicity and strength, where strength is understood in terms of success of PEC-inferences. It would be interesting to investigate this analogy in more detail, to see whether problems facing the best-system view of laws carry over to the epistemic theory of causality.

This brings us to the question of realism. The epistemic account of causality in terms of causal beliefs is in principle compatible with either anti-realism or realism about causality (Williamson, 2005, §9.4), since the claim that causality is best understood epistemically leaves open whether causal relationships are really 'out there', in the sense of being characterisable without invoking epistemological considerations. The question then arises as to which is the right stance to take. The anti-realist view may be motivated by disillusionment with current realist analyses of causality, which seem to fall to counterexamples and problems accounting for the epistemology of causality. On the other hand, realism about causality might be motivated by the thought that the existence of a non-epistemic causal relation would offer the best explanation of the utility of our causal beliefs. The anti-realist can undermine the motivation behind realism, however, by arguing that it is the fact that causal claims tend to tally with underlying mechanisms that accounts for their explanatory utility, while it is the fact that causal claims tend to coincide with difference-making relationships that accounts for their utility with respect to prediction and control.

The realist might try a different line of attack here by appealing to the aforementioned connection between the epistemic theory and a best-system account of causality, since best-system accounts in general seem to be amenable to realism. Suppose that a best-system account of causality can indeed be derived from the epistemic account: A is a cause of B iff that causal relationship is included in some system of causal claims that best balances simplicity against strength, where strength is understood in terms of success of PEC-inferences. The realist might point out that causality then appears to be a worldly relation, independent of epistemological concerns. The anti-realist might counter that judgement and reason enter into the best-system account in various ways. First, the strength of a system of causal claims is understood epistemologically, in terms of successful inferences. Second, epistemological concerns may influence standards of simplicity and balance. In response to this worry, Lewis (1980, p. 123) maintains that, given fixed standards, laws (in our case, causal laws) need not depend on us in any way. However, if there is disagreement as to standards of simplicity and balance then as to which standard is to be preferred may indeed depend on our epistemological concerns. Third, Lewis' best-system view of laws invokes the notion of overall similarity of possible worlds; epistemological considerations may help determine what constitutes overall similarity (Lewis, 1986b). Fourth, other best-system accounts relativise causal claims to natural kinds; epistemological considerations may be required to select an appropriate set of kinds, and hence to determine whether A causes B simpliciter (Cohen and Callender, 2009, §4.3). Given these worries, it is by no means clear yet that some such best-system account of causality can underwrite realism about causality.

Failing that line of attack, the realist might try another way out. The realist might point out that, even if the arguments of this paper are accepted, there is still room for a characterisation of causality that does not appeal to epistemological

considerations. The realist can accept that difference-making analyses of causality fail because of problems to do with overdetermination, but insist on a necessary condition on causality: if E is not overdetermined and C is a cause of E then there is a chain of difference making from C to E. The realist can accept that mechanistic analyses of causality fail because of problems to do with absences (and perhaps also in other cases such as double prevention), but insist on a necessary condition on causality: if C and E are not absences (and there is no double prevention) and C is a cause of E then there is a chain of mechanisms from C to E. The realist can then say that *the* causal relation is the smallest relation that satisfies these necessary conditions, perhaps in conjunction with other necessary conditions.

If some such characterisation of the causal relation were workable, that would of course not tell against the epistemic account of causality. Arguably, the epistemic account would still be required in order to explain why this set of necessary conditions provided a characterisation of the *causal* relation, as opposed to some ad hoc relation. An answer would proceed along the following lines: it is its relationship to PEC-inferences that provides the primary characterisation of causality; if a necessary-condition characterisation were also workable that would be because those conditions were conditions for successful PEC-inference.

In sum, the epistemic account of causality leaves open the question of whether realism or anti-realism about causality is most credible, and there is a range of ways in which one might fruitfully attempt to answer this question.

Acknowledgements

This research was conducted as a part of the research project *Mechanisms and the evidence hierarchy*, supported by the UK Arts and Humanities Research Council. I am very grateful to Lorenzo Casini, Start Glennan, Julien Murzi, Erik Weber and an anonymous referee for comments on earlier versions of this paper.

Bibliography

Anscombe, G. (1971). *Causality and determinism.* Cambridge University Press.

Bechtel, W. and Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36:421–441.

Bradford Hill, A. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58:295–300.

Cartwright, N. (2004). Causation: one word, many things. *Philosophy of Science*, 71:805–819.

Cartwright, N. and Efstathiou, S. (2011). Hunting causes and using them: Is there no bridge from here to there? *International Studies in the Philosophy of Science*, 25(3):223–241.

Chadeau-Hyam, M., Athersuch, T. J., Keun, H. C., Iorio, M. D., Ebbels, T. M., Jenab, M., Sacerdote, C., Bruce, S. J., Holmes, E., and Vineis, P. (2011). Meeting-in-the-middle using metabolic profiling—a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers*, 16(1):83–88.

Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2012). The evidence that evidence-based medicine omits. *Preventative Medicine*, DOI 10.1016/j.ypmed.2012.10.020.

Cohen, J. and Callender, C. (2009). A better best system account of lawhood. *Philosophical Studies*, 145:1–34.

Craver, C. F. (2007). *Explaining the brain.* Oxford University Press.

Darden, L. (2013). Mechanisms versus causes in biology and medicine. In Chao, H.-K., Chen, S.-T., and Millstein, R., editors, *Mechanism and Causality in Biology and Economics*, Theory, Philosophy, and History of the Life Sciences. Springer, Berlin. Chapter 2.

Dowe, P. (2000). *Physical causation.* Cambridge University Press, Cambridge.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69:S342–S353.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44:49–71.

Hall, N. (2004). Two concepts of causation. In Collins, J., Hall, N., and Paul, L., editors, *Causation and counterfactuals*, pages 225–276. MIT Press, Cambridge MA and London.

Illari, P. M. and Williamson, J. (2011). Mechanisms are real and local. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 818–844. Oxford University Press, Oxford.

Illari, P. M. and Williamson, J. (2012). What is a mechanism? thinking about mechanisms *across* the sciences. *European Journal for Philosophy of Science*, 2:119–135.

Lakatos, I. (1978). Understanding Toulmin. In Worrall, J. and Currie, G., editors, *Mathematics, science and epistemology: philosophical papers*, volume 2, pages 224–243. Cambridge University Press, Cambridge.

Lewis, D. K. (1973). Causation. In *Philosophical papers*, volume 2, pages 159–213. Oxford University Press (1986), Oxford.

Lewis, D. K. (1980). A subjectivist's guide to objective chance. In *Philosophical papers*, volume 2, pages 83–132. Oxford University Press (1986), Oxford. With postscripts.

Lewis, D. K. (1986a). Causal explanation. In *Philosophical papers*, volume 2, pages 214–240. Oxford University Press (1986), Oxford.

Lewis, D. K. (1986b). Counterfactual dependence and time's arrow. In *Philosophical papers*, volume 2, pages 32–66. Oxford University Press. With postscripts.

Longworth, F. (2006a). *Causation, counterfactual dependence and pluralism.* PhD thesis, University of Pittsburgh.

Longworth, F. (2006b). Causation, pluralism and responsibility. *Philosophica*, 77:45–68.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67:1–25.

Menzies, P. and Price, H. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science*, 44:187–203.

Price, H. (1991). Agency and probabilistic causality. *British Journal for the Philosophy of Science*, 42:157–176.

Psillos, S. (2009). Causal pluralism. In Vanderbeeken, R. and D'Hooghe, B., editors, *Worldviews, Science and Us: Studies of Analytical Metaphysics. A Selection of Topics From a Methodological Perspective*, pages 131–151, Singapore. World Scientific.

Reichenbach, H. (1956). *The direction of time.* University of California Press, Berkeley and Los Angeles, 1971 edition.

Reiss, J. (2011). Third time's a charm: causation, science and Wittgensteinian pluralism. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 907–927. Oxford University Press, Oxford.

Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2):157–170.

Russo, F. and Williamson, J. (2011). Epistemic causality and evidence-based medicine. *History and Philosophy of the Life Sciences*, 33(4):563–582.

Russo, F. and Williamson, J. (2012). EnviroGenomarkers: the interplay between mechanisms and difference making in establishing causal claims. *Medicine Studies: International Journal for the History, Philosophy and Ethics of Medicine & Allied Sciences*, 3:249–262.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton University Press, Princeton NJ.

Salmon, W. C. (1998). *Causality and explanation.* Oxford University Press, Oxford.

Weinberg, R. A. (2007). *The biology of cancer.* Garland Science, New York and Abingdon.

Williamson, J. (2005). *Bayesian nets and causality: philosophical and computational foundations.* Oxford University Press, Oxford.

Williamson, J. (2006). Causal pluralism versus epistemic causality. *Philosophica*, 77:69–96.

Williamson, J. (2009). Probabilistic theories. In Beebee, H., Hitchcock, C., and Menzies, P., editors, *The Oxford Handbook of Causation*, pages 185–212. Oxford University Press, Oxford.

Williamson, J. (2010). *In defence of objective Bayesianism.* Oxford University Press, Oxford.

Williamson, J. (2011). Mechanistic theories of causality. *Philosophy Compass*, 6(6):421–447.

Woodward, J. (2003). *Making things happen: a theory of causal explanation.* Oxford University Press, Oxford.