·40·
The causal-process-model theory of mechanisms

Phil Dowe

Abstract

Wesley Salmon and I have argued that causation and causal explanation needs to appeal to causal processes understood in terms of conserved quantities. This has the consequence of ruling about absence 'causation' as being genuine causation. Carl Craver has argued persuasively that absences are crucial in causal explanations in neuroscience, and so he gives an account of mechanisms in terms of causal relevance where the latter is understood along the lines of causal modelling. This allows for absences to be causes and hence to feature in causal explanations, but it is not compatible with the claim that causal explanation needs to appeal to causal processes understood in terms of conserved quantities. I therefore offer an account of mechanisms, in particular the role of causal relevance in mechanisms, which can respect the theory that causation involves causal processes understood in terms of conserved quantities, but which also allows absences to figure in causal explanation.

40.1 Introduction

In some areas of science, the right correlations are routinely taken to indicate causality. In other areas, it's common enough that a scientist might be unwilling to infer causality from correlations without also knowing *how* one factor is responsible for another. Some philosophers take this to show that correlations are insufficient for causation. They might alos take this to indicate that scientific explanation, in those areas of science at least, involves appeal to mechanisms. I agree, for some areas of science, that this indicates both that correlations are insufficient for causation and that mechanisms are the basis for scientific explanation.

In what follows I survey two attempts to give an account of mechanistic explanation, namely those due to Salmon (1984) and Craver (2007). The question to be addressed is: what theory of mechanisms can account for the idea that scientists sometimes seek mechanisms that underlie correlations. I will assume that the answer cannot be 'more correlations'.

978–0–19–957413–1 40-Mckay-Illari-c40-drv Mckay (Typeset by SPi, Chennai) 866 of 928 November 4, 2010 17:5
OUP UNCORRECTED PROOF – FIRST PROOF, 4/11/2010, SPi

40.2 Salmon's mechanistic theory of explanation

Wesley Salmon draws the following lesson, notably in *Scientific Explanation and the Causal Structure of the World* (1984). There are two tiers to scientific explanation. The first tier involves set of statistical relevance relations. But this is not enough to guarantee causation. We need a second tier, which involves exhibiting the causal connections. Together, this provides an account of explanation which is 'ontic' rather than epistemic. Hempel's account of explanation as arguments, for example, makes explanation an epistemic matter, but Salmon's account in terms of objective facts of statistical relevance and causal connection makes for an ontic account.

According to Salmon's Statistical Relevance Criterion, explanations involve first an assemblage of facts, facts which are statistically relevant to the thing to be explained. An event C is statistically relevant to another event E if the probability of E is affected by whether or not C occurs: C is statistically relevant to E iff $\mathrm{P}(E|C) \neq \mathrm{P}(E)$ and C is statistically irrelevant to E iff $\mathrm{P}(E|C) = \mathrm{P}(E)$. For Salmon these probabilities are to be understood as relative frequencies.

An explanation of an event E involves four steps. First, we begin with the prior probability of E – the likelihood of that event relative to an appropriate reference class (R). This will be of the form $\mathrm{P}(E|R)$. Second, we need to find relevant partitions of this reference class. A partition is relevant if the probability of E is different in the relevant cells (parts of the reference class). For example, if the probability that a female will die of lung cancer is the same as the probability that a male will, the male/female partition is not relevant to death by lung cancer. However, whether an individual smoked or not will be relevant if the probability of the individual getting cancer if they smoke is different to the probability that they get cancer if they don't smoke. Third, we need to know the posterior probabilities, that is, the probabilities in each cell after we make the partition. We need to know the probability of getting cancer if one smokes, and if one doesn't smoke; ie $\mathrm{P}(E|S)$ and $\mathrm{P}(E| \sim S)$. Finally, we need to locate the individual in one of these cells. Then, if we are satisfied that there are no further relevant partitions, the explanation involves citing all the factors in the definition of that cell.

Salmon gives a fictional example of an American teenager Albert who is convicted of the offence (O) of stealing a car. Albert is a male from an urban environment. Take the class of American teenagers (T) as the reference class. Dividing it into male (M) and female (F) happens to be a relevant partition (it's more likely that teenage males will commit offences in America than female teenagers) and so does dividing into urban (U) and rural (R) background (city bred teenagers are more likely to commit offences than those who are rural bred). This gives four cells in the reference class: male-urban, male-rural, female-urban, female-rural. The probabilities for each will be different. We select the relevant cell: male-urban, and the fact that the probability in that

cell is higher than the prior probability – provided there are no further relevant partitions – explains the teenager's being convicted. One of the advantage of this approach as Salmon sees it is that it allows for low probability explanations. Explanatory factors raise the probability of the thing they explain, but don't necessarily make it probable.

Exhibiting the statistically relevant facts is the first step in explanation, but this needs to be added to, by exhibiting the causal links between the fact to be explained and the statistically relevant facts. When we have statistically relevant facts linked by a causal process to the thing to be explained, then we have a satisfactory explanation. Salmon therefore turns his attention to the question of causality. His contribution here is twofold: he offers a persistent line of criticism against the probabilistic theories of causation, and he offers his own positive account.

The essence of the probabilistic theory of causality is the idea that a cause raises the probability of its effect (Suppes 1970). Salmon's argument against this theory concerns counterexamples where a particular causal chain contains elements that do not stand in the probability raising relation. In an example due originally to Deborah Rosen a golfer slices her shot, but by sheer fortune hits a tree branch, and the ball bounces back onto the green and into the hole. The slice lowered the chance of a hole in one, but in fact caused it.

Salmon's positive account treats causation primarily as a property of individual processes. Salmon proposes to overcome traditional difficulties with determining the nature of the causal relation by treating causation as primarily a characteristic of continuous processes rather than as a relation between events. This treatment involves two elements, the production and the propagation of causal influence. The latter is achieved by causal processes. Salmon's views on how to characterize causal processes underwent various changes which are not relevant to our purposes. In his 1997 *Philosophy of Science* paper Salmon presents the following revised theory of causality:

> A causal process is the world line of an object that transmits a non-zero amount of a conserved quantity at each moment of its history (each spacetime point of its trajectory) (Salmon 1997, p. 468)

The concept of transmission is to be understood by the following definition:

> A process transmits a conserved quantity between A and B ($A \neq B$) if and only if it possesses [a fixed amount of] this quantity at A and at B and at every stage of the process between A and B without any interaction in the open interval (A, B) that involve an exchange of that particular conserved quantity (Salmon 1997, p. 463)

Thus, Salmon appeals to a special kind of regularity which involves the possession of a fixed amount of a conserved quantity at every spacetime point of the process (for my attempt to do the same, in the context of the causal theory of explanation see Dowe 1992).

So a causal process is a worldline of an object that possesses a conserved quantity, a causal interaction is an intersection of worldlines involving exchange of conserved quantities. For one event to stand in a causal relation to another they must be connected by a set of causal processes and interactions. This, however, gives us a necessary condition, but not a sufficient condition. We also need to say in what respects an object at one time is causally relevant to an object at another time. (For a discussion of my attempt to satisfy this requirement, and objections to that see Dowe, 2007.) However, that it is only a necessary condition won't matter in this context because Salmon conjoins the causal processes-interactions requirement with the statistical relevance requirement, which should rule out irrelevances.

What will matter is the problem of chance lowering causes, as identified by Salmon himself, and used by him to rule out the chance-raising theory of causation. While it is true that a process theory of causation will not be open to an objection from chance-lowering causes, Salmon's theory of explanation will be, because it conjoins the process theory with a statistical relevance requirement. This means we cannot appeal to chance-lowering causes to explain their effects.

In any case, Salmon's account doesn't go far enough to give us an account of explanations or mechanisms. First, the account seems geared to explaining particular events, whereas mechanisms provide general explanations. Explaining why Albert was convicted no doubt needs to establish that he was an actual cause of the offense, and this may well be done in terms of causal processes. But a mechanism of the sort appealed to for example in medical sciences would provide a general explanation.

Take the following example from lipid metabolism research:

> The association between abdominal fat accumulation and risk of chronic diseases, including type II diabetes and coronary heart disease, has long been recognized. Insulin resistance may be a key factor in this link. Many studies have pointed to an association between insulin resistance and intra-abdominal fat accumulation (visceral obesity). However there is no clear proof of a causal link between visceral fat accumulation and insulin resistance. In assessing the probability of a causal link, it is useful to consider potential mechanisms. One such potential causal link is the release of non-esterified fatty acids from visceral fat into the portal vein, so that they have direct effects on hepatic metabolism. Visceral fat has been shown in many studies to exhibit a high rate of lipolysis compared with subcutaneous fat depots. However, if the idea that visceral fat releases fatty acids into the portal vein at a high rate is examined critically, a number of difficulties appear. Not least of these is the fact that continued high rates of lipolysis should lead to the disappearance of the visceral fat depot, unless these high rates of fat mobilization are matched by high rates of fat deposition. There is far less evidence for high rates of fat deposition in visceral adipose tissue, and some contrary evidence. Evidence for high rates of visceral lipolysis in vivo from studies involving catheterization of the portal vein is not strong. If this potential link is discounted, then other reasons for the relationship between visceral fat and insulin resistance must be

considered. One is that there is no direct causal link, but both co-correlate with some other variable. A possibility is that this other variable is subcutaneous abdominal fat, which usually outweighs intra-abdominal fat several-fold. Subcutaneous fat probably plays the major role in determining systemic plasma non-esteri?ed fatty acid concentrations, which are relevant in determining insulin resistance. (Frayn 2000, p. S71)

Note that correlations by themselves are open to at least two causal interpretations, a direct causal connection and a common causal connection: two factors are the effects of a common cause. There are of course plenty of techniques to distinguish the two interpretations in terms of correlations. But these are neither conclusive, nor representative of how scientists in certain areas approach the question. Scientists rather seek mechanisms, as Salmon urges.

But in this example, we want a mechanism to explain insulin resistance in general in the first instance, not an explanation of what occurred in one particular instance. Mechanisms explain how things work, and this entails saying how various *alternatives* operate. Mechanisms typically have more than one possible value of an input. If we are, for example, to appeal to the effects of non-esteri?ed fatty acids in the portal vein on hepatic metabolism, we cannot simply trace the connections between single values of the relevant variables.

And finally, to explain how something works we sometimes need to account for the fact that something of interest didn't happen (Woodward 2003, pp. 227ff). For example, the idea that high rates of lipolysis lead to the *disappearance* of the visceral fat depot, not to mention the notion of insulin resistance itself, seems to involve absences rather than positive occurrences (see below for more examples). But there's no causation by or of absences according to the causal process theory, so there must be more to a mechanism than just an actual set of actual causal processes and interactions.

40.3 Craver's account of mechanisms

Carl Craver's *Explaining the Brain: Mechanisms and the Mosaic Unity of Science* (2007) provides an account which does each of these things, with a specific orientation to neuroscience. According to Craver a mechanism is a set of entities with associated productive activities organised so as to constitute some phenomenon. 'Activities' is a 'filler term' for a set of causal components, meaning that to count as a mechanism the entities that constitute the phenomenon must be linked by "causal relevance". To count as a mechanism the entities must also be linked with the phenomenon they constitute by 'constitutive relevance', which in turn is understood in terms of mutual manipulability: change the phenomenon and you change the parts, change a part and you change the phenomenon.

Drawing on Salmon's distinction between 'ontic' and 'epistemic' explanation, a distinction to which he subscribes, Craver notes an ambiguity in the word 'explanation':

> Sometimes explanations are texts – descriptions, models, or representations of any sort that are used to convey information from one person to another. Explanatory texts are the kinds of things that are spoken and written, and drawn. They are the kinds of things that can be more or less complete and more or less accurate. They are representations. Other times explanations are objective portions of the causal structure of the world, the set of factors that bring about or sustain a phenomenon (call them objective explanations). What explains the accident? The ice on the road, the whiskey, the argument, the tears, and the severed break cables. What explains the release of neurotransmitters? The action potential, Ca^{2+} influx, vesicular binding, and fusion. There are mechanisms (the objective explanations) and there are their descriptions (explanatory texts). Objective explanations are not texts; they are full-bodied things. They are facts, not representations. They are the kinds of things that are discovered and described. There is no question of objective explanations being 'right' or 'wrong,' or 'good' or 'bad'. They just are. Craver 2007, p. 27)

Causes and mechanisms, then, are "things in the world". A second implication of explanations being ontic – the first is that they aren't arguments – is that they aren't representations.

According to Craver mechanisms are typically multi-level, requiring at the level of description what he calls the 'mosaic'. This includes multi-level causation and in particular the fact that higher levels can be causal. As he sees it this ties in with integrative neuroscience, where mechanisms, and hence explanations, involve multiple levels and cross multiple fields so that entities dealt with in different fields jointly constitute explanations. This notable aspect of Craver's account will not be dealt with in this chapter.

To give an account of causal relevance Craver draws heavily on Woodward (2003): a 'variable X is a cause of variable Y in conditions W, if and only if it is possible in conditions W to change the value of Y with an ideal intervention that changes the value of X' (Craver 2007, p. 94), where an *ideal* intervention I on X with respect to Y is a change in the value of X that changes Y, if at all, *only via* the change in X. More specifically, this requirement implies that:

(I1) I does not change Y directly;

(I2) I does not change the value of some causal intermediate S between X and Yexcept by changing the value of X;

(I3) I is not correlated with some other variable M that is a cause of Y; and

(I4) I acts as a 'switch' that controls the value of X irrespective of X's other causes, U. (Craver 2007, p. 96)

Craver provides a number of detailed examples. One (Craver 2007, pp, 65–72) is a certain type of 'Long-Term Potentiation' (LTP), where

changes to a synapse are produced which are thought to underlie certain kinds of learning. Two neurons are involved, the pre-synaptic and post-synaptic. Firing of the pre-synaptic neuron releases the neurotransmitter glutamate, which crosses the synapse and binds to the receptors on the post-synaptic cell. There are two types of receptors relevant here, NMDA receptors (because they are responsive to N-methyl-D- aspartate) and AMPA receptors (because they are responsive to α-amino-3-hydroxyl-5-methyl-4-isoxasolepropionic acid). The function of the NMDA receptor is to create a Ca^{2+} selective channel. However, if the post-synaptic cell is polarized, the Ca^{2+} channel is blocked by large Mg^{2+} ions. The function of the AMPA receptor is to depolarize the post-synaptic cell and thereby repel the Mg^{2+} ions from the channel, enabling Ca^{2+} to flow through the MNDA receptor. The increase in Ca^{2+} concentrations in the post-synaptic neuron in turn activates a number of intracellular biochemical pathways including those responsible for the production of proteins used to alter the structure of the synapse. The firing of the pre-synaptic neuron can be induced experimentally, and this intervention is correlated with changes in synaptic efficiency (see Craver 2007, p. 67). This is already an impressive amount of detail, and I've only summarized the bare bones of Craver's example.

At first pass, the idea of singular productive activities looks like it might provide the right 'ontic something' that underlies and explains correlations. But there are worries, ironically, about both whether the account really is ontic in the relevant sense, and also whether it can be said to provide that something beyond correlations which explain correlations. Both concerns focus on the Woodwardian explication of causal relevance. It's true that there's more to Craver's mechanisms than causal relevance – they also exhibit constitutive relevance (although arguably the same worries re-emerge) and organization. If the concerns have merit then it would seem the latter must be what really does the work. I don't want to sound too pessimistic about the prospects of its doing so, as there is much to commend the account in Craver's book including the detailed workings of examples. My point is that the account of causal relevance *per se* raises these concerns. And it is apparent that Craver intends the notion causal relevance to indeed carry the burden (i.e. to be that ontic something which explains correlations). He says:

> In saying that activities are *productive*, I mean that they are not mere correlations . . . and most fundamentally, that they can potentially be exploited for the purposes of manipulation and control. (Craver 2007, p. 6)

> More explicitly: what it means to say that one stage of a mechanism is productive of another (as I suggest in Machamer *et al*. 2000; Craver and Darden 2001; Darden 2002), and to say that one item (activity, entity, or property) is relevant to another, is to say, at least in part, that one has the ability to manipulate one item by intervening to change another. (Craver 2007, pp. 93–4)

978–0–19–957413–1 40-Mckay-Illari-c40-drv Mckay (Typeset by SPi, Chennai) 872 of 928 November 4, 2010 17:5
OUP UNCORRECTED PROOF – FIRST PROOF, 4/11/2010, SPi

Talk of 'one having the ability' should not be taken too literally. As for Woodward, this is a way of appealing to certain counterfactuals, as outlined above. However, as there is no account of the truth conditions of the counterfactuals, but rather an appeal to experimentation to discern their truth, Craver is open to the charge that the account of causal relevance and hence mechanisms cannot provide that something beyond correlations. It is simply a particular subset of correlations (not the 'mere' correlations) that the account rests on. I say 'might be open to the concern' because, given the absence of an account of causation *per se*, and indeed of the truth conditions for the counterfactuals, the theory is sufficiently incomplete to allow that it may be elaborated in such a way that obviates the concern.

The second concern, again tied up with the appeal to causal modelling, is that the account seems to make mechanisms epistemic rather than ontic, contrary to Craver's stated aim. The causal modelling approach to causation, and the manipulability approach derived from (or inherent in) it, involves models which are *abstractions* (see Menzies 2004, pp. 154–7). There are many ways to abstract, and thus, as is well known, causal modeling makes causation (here causal relevance) model relative. This in itself makes the notion of causal relevance – a key component in a mechanism on Craver's account – look rather more like a representation than an objective feature of the world.

But whether that is so or not, it's clear that we never will get beyond correlations on this account. To illustrate these concerns, let's return to Craver's example. We could simply model some of the phenomena in LTP by the following model:

$S = 1$ if the pre-synaptic neuron is stimulated, S=0 if not.

$R = 1$ if there is an increase in post-synaptic response time, R=0 if not.

$\Pr(R = 1|S = 1) > \Pr(R = 1|S = 0)$.

S is causally relevant to R by the definition: an intervention on S raises the probability of $R = 1$, and only does so via S. But all we've done is encode the correlations that we should explain in an adequate explanation. So let's give a slightly more detailed model:

$S = 1$ if the pre-synaptic neuron is stimulated, $S = 0$ if not.

$C = 1$ if the concentration of Ca^{2+} in the post-synaptic neuron are above a certain threshold, $C = 0$ if not.

$M = 1$ if Mg^{2+} ions block the Ca^{2+} channel, $M = 0$ if not.

$\Pr(C = 1|S = 1) > \Pr(C = 1|S = 0)$.

$\Pr(M = 0|S = 1) > \Pr(M = 0|S = 0)$.

$\Pr(C = 1|M = 1) = 0$.

$\Pr(C = 1|M = 0) \neq 0$.

Causal modelling would represent this on a two-pathway graph with an arrow from S to C, an arrow from S to M and an arrow from M to C. An intervention on S would change the values of M and C, and an intervention on M would change the value of C. This nicely captures the way in which experiments inform our understanding of causation and mechanisms. But, again, all we have done is encode more correlations. This is inadequate if our task is figuring out what it means to move beyond correlations to a mechanism.

I suggest that we rather appeal to causal processes to begin to capture the force of Craver's impressive examples. What carries the argument is the appeal to processes such as the movement of neurotransmitters, and the flow of Ca^{2+} ions. These processes qualify as causal processes on the account set out above because glutamate and Ca^{2+} ions have conserved quantities like mass and charge. Charge in particular is most pertinent to explanations in neuroscience. Craver objects that the conserved quantity theory drives us to fundamentalism: 'CQ also presents a view of causation tailor-made for physicalist/fundamentalist metaphysics. If causal interactions are exchanges of conserved quantities, and if conserved quantities are found only at the fundamental level, then all causation is located at the fundamental level.' (Craver 2007, p. 77). However, this misreads the account set out above. The CQ theory appeals to the trajectories of objects possessing conserved quantities. A steel ball possesses mass and charge as much as an electron does, and so too for a Ca^{2+} ion.

However, we still have the problem of absence causation. In the example just discussed, both $S = 1$ and $M = 0$ cause $C = 1$. $M = 0$ is the absence of Mg^{2+} ions in the Ca^{2+} channel, and Craver makes a convincing case for admitting absences such as this one into explanatory mechanisms in neuroscience (e.g. 2007, pp. 80–81). I think this has to be accepted. But absences are not causes and the CQ theory rules them out as causes. But the problem with absence causation is not just that the CQ theory rules them out. On a counterfactual theory of causation (for example Lewis 2004), absence causation violates relativity (Dowe 2009). (Ironically, Craver's initial characterization of a mechanism as involving singular productive mechanisms looks like it would rule out absences: how can something's not occurring be 'productive'.)

Nevertheless, Craver also runs into trouble trying to say exactly how absences figure in explanations. Or more specifically, why they don't when they don't. The problem concerns what Peter Menzies calls 'profligate causation' (Menzies 2004, p. 145). If absences are causes then there are far more causes than we expect intuitively. Craver notes

> A … problem raised against the acceptance of negative causes is that there are too many of them, and most negative causes are of no use for understanding explanation in neuroscience. As Dowe (2004) and Beebee (2004) argue, many instances of negative causation run counter to our common sense, scientific, and theoretical uses of the

978–0–19–957413–1 40-Mckay-Illari-c40-drv Mckay (Typeset by SPi, Chennai) 874 of 928 November 4, 2010 17:5
OUP UNCORRECTED PROOF – FIRST PROOF, 4/11/2010, SPi

concept of 'cause', and there is no available account of negative causation that allows in all and only the intuitively satisfactory instances. (Craver 2007, p. 83)

It would seem, then, that a satisfactory account of mechanistic explanation should tell us when we can and cannot appeal to absence causation. Craver's response is this:

> The extravagant cases of negative causation can be handled in a number of ways. Some negative causes are too improbable or abnormal to be included in explanatory texts or even counted as causes. Others are ruled out by, for example, legal, moral, and epistemic factors that determine the salience of a fact in a particular discussion (see Beebee 2004). . . . Consider a neuroscientific example: is the gasoline in my car's tank a cause of the instance of LTP in the Petri dish? It is likely true that if I had doused the dish with the gasoline, then the cells would not induce LTP, but it seems odd to think of the absence of gasoline as a cause of LTP. Although I do not have a general formula for ruling out non-explanatory causes of this sort, it is clear enough that gasoline is neither normally part of cells nor part of their extracellular environment. Gasoline is not part of the set-up or background conditions under which the cell normally operates. It is not a cellular constituent. Gasoline levels do not vary as the mechanism works. The distinction between intuitive and counterintuitive cases is a psychological distinction that is drawn on a number of different grounds in different epistemic contexts. (Craver 2007, p. 85)

This appears to make what counts as a mechanisms dependent on an individual's or a group's psychology. But perhaps a better interpretation is to take the above response to apply not to explanations but to explanatory texts, and objective explanations just do involve numerous absences, all of which are actually causally relevant. Given that Craver refers here to 'normal operation', certain approaches (Menzies 2004, Hitchcock 2007), which attempt to deal with the problem of profligate absence causation by appeal to a notion of normal operation might be of use to Craver here, if one could establish an objective notion of 'normal operation' for neuroscience. However, this is not the direction I want to take.

40.4 Causal process mechanisms

Drawing together some of the lessons from our discussion of Salmon's and Craver's accounts, we shall require the following desiderata of an account of a mechanism:

1. Mechanisms should explain correlations, setting out for example how one correlate causes the other.
2. Mechanisms should encode alternatives.
3. Mechanisms should include absences.
4. Mechanistic explanation should be ontic not epistemic.

A mechanism is not necessarily named as a phenomenon, but often is when used to explain. A mechanism can but need not involve sub-mechanisms. Mechanisms are glued together by causation.

We start with a model of a type of situation. We have a choice of the variables, *U, V, W,* ... and a choice of a partition of each variable into incompatible and jointly exhaustive values: u_1, u_2, u_3, v_1, v_2, etc. There are laws which apply to this kind of situation. Based on the laws, for some set of values of initial variables (yet to be defined) each pair of values of a variable pair, either will or won't be connected in the right way by causal processes and interactions. (This account would also work with other physical connection theories, such as that of Fair 1979, but for my reasons for rejecting that account see Dowe 1995. For an attempt to combine causal modeling and causal processes in an account of causation, see Handfield *et al.* 2008.)

To generate a *Causal Process Model* we write down all values that are thus connected, with an arrow indicating the connections:

$u_i \rightarrow v_j$

$v_j \rightarrow w_k$

etc.

When two variable values are connected by such an arrow, the antecedent is a cause and the consequent is an effect (the process theory of causation). The Relevance Condition on a mechanism requires that some value of every variable in the model is connected one way or another to some value of every other variable. Where a causal connection depends on some variable value which is not a cause of the variable value named as the antecedent (compare Hitchcock 2001), write down the required value before the antecedent, say:

$w_k, u_i \rightarrow v_j$.

The motivation for this technique is to make transparent what conditions would interrupt the causal connection. We remove indirect connections by the following rule:

Non redundancy: Remove any connection in the set that can be generated by Transitivity from other connections in the set.

Transitivity: If $u_i \rightarrow v_j$ and $v_j \rightarrow w_k$ then $u_i \rightarrow w_k$.

This minimal set of connections together with the variable value set constitutes the causal process model. (Whether causation is transitive is a controversial matter. For the case that it is see Lewis 2004; for the case that it isn't see Hitchcock 2001. I side with the former, but won't argue for that here.)

Take as an example the standard late preemption case. Billy and Suzi throw rocks at a glass bottle, Suzi's throw is slightly stronger and will smash the

978–0–19–957413–1 40-Mckay-Illari-c40-drv Mckay (Typeset by SPi, Chennai) 876 of 928 November 4, 2010 17:5
OUP UNCORRECTED PROOF – FIRST PROOF, 4/11/2010, SPi

glass, but were Suzi not to throw, Billy's throw would smash the glass. Let $B(1,0)$ stand for Billy's throwing or not, $S(1,0)$ for Suzi's throwing or not, and $G(1,0)$ the glass bottle being smashed or not. We model the following causal connections:

$s_1 \rightarrow g_1$
$s_0, b_1 \rightarrow g_1$.

This is not a case that we would normally call a mechanism, but the analysis will apply to any simple backup mechanism where a signal S and its back-up B both typically fire, and when they both do S causes effect G, but when S doesn't fire, B causes H. The mechanism represented by the Causal Process Model enables us to explain various outcomes, depending on the input values of S and B. We then understand how the mechanism works in general, because we know the possible causal processes, and how their operation depends, or not, on other variables. So mechanisms explain in both the particular and general sense. They explain particular events, and they explain how in general a system works. In the particular case, the explanation can appeals to actual causal connections that connect the actual values, and possible connections that connect non-actual values. In the general case, explanations appeal to possible connections between possible values.

Any thus defined mechanism will be an approximation in an important sense. If a mechanism is a system of connections that would hold given certain values of certain variables, then there will always be another more detailed mechanism which the first mechanism approximates. That the system of connections would obtain given certain values of certain variables holds only on the assumption that other interfering factors – factors not modelled – are not present. A more detailed mechanism would include some of these factors. In practice no model would be so completely detailed as to avoid such an assumption. As Mill pointed out, 'a special enumeration of ... the negative conditions ... of any phenomenon ... would generally be very prolix' (Mill 1843, pp. 370–1). Nevertheless mechanisms explain correlations between the variables in the mechanism. We are interested in certain stable correlations which arise because other possible interfering factors are held fixed; perhaps because they are rare or easily controlled in an experimental situation.

On this account explaining by appeal to mechanisms is an ontic matter. Actual causal connections are things that are 'in the world' no matter how they are represented, and possible causal connections are guaranteed by the laws. Is this claim threatened by the fact that it is model (mechanism) relative which interfering factors are included? No. First, whether a particular actual causal connection obtains or not is not model relative. It's true that what causal connections one appeals to in order to explain something depends on the

choice of model; i.e. on which mechanism one appeals to. That doesn't mean the explanation is not ontic. Second, it is model relative which interfering factors a non-actual causal connect depend on. But again, this does not mean the account is not ontic, just that different mechanisms specify differently – more or less completely – the conditions under which the connection would hold.

The discussion of interfering factors leads us into the question of how absences enter into mechanisms and explanations. Take our model of LTP

$S = 1$ if the pre-synaptic neuron is stimulated, $S = 0$ if not.

$C = 1$ if the concentration of Ca^{2+} in the post-synaptic neuron is above a certain threshold, $C = 0$ if not.

$O = 1$ if the Ca^{2+} channel is open, $O = 0$ if not.

$M = 1$ if Mg^{2+} ions block the Ca^{2+} channel, $M = 0$ if not.

$P = 1$ if the post-synaptic cell is polarized, $P = 0$ if not.

$N = 1$ if there is sufficient Ca^{2+}outside the post-synaptic cell, $N = 0$ if not A causal process model is then

$S(1,0)$, $C(1,0)$, $O(1,0)$, $M(1,0)$, $P(1,0)$, $N(1,0)$ and

$s_1 \rightarrow o_1$

$s_1 \rightarrow p_0$

$o_1, p_1 \rightarrow m_1$

$m_{0,}, o_1, n_1 \rightarrow c_1$.

The opening of the Ca^{2+} channel, and the absence of Mg^{2+} ions each allow the Ca^{2+} flow, but do not cause it. Nevertheless, they are part of the mechanism, and enter into the explanation. (Compare my account of prevention and omission, originally in Dowe 1999.) So, in a particular case the absence of Mg^{2+} ions might help explain the Ca^{2+} concentration (the explanation appeals to an omission). In another case the presence of Mg^{2+} ions explains the absence of adequate Ca^{2+} concentration (the explanation appeals to a prevention).

What about the problem of profligate omissions? Craver's position is that some absences enter into mechanisms, others do not. The absence of gasoline in the Petri dish is not part of the mechanism of LTP. The problem then is that the definition of the *mechanism* becomes interest relative. On my account there are two mechanisms, one that includes the absence of gasoline, and another, an approximation to the first, which omits that interfering factor. Gasoline floods don't normally occur in the brain, and are controlled for in experimental situations, so the second mechanism is what we appeal to in neuroscience. That mechanism explains the correlations that obtain because of the general absence of that interfering factor. But that certain causal con-

nections in that mechanism would occur under certain conditions is true only because we assume the absence of that interfering factor.

References

Beebee, H. (2004), Causing and nothingness, in J. Collins, N. Hall and L. Paul (eds.), *Causation and Counterfactuals,* Cambridge, Mass.: MIT Press, pp. 291-308.

Craver, C. (2007), *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience,* Oxford: Oxford University Press.

Dowe, P. (1992), An empiricist defence of the causal account of explanation, *International Studies in the Philosophy of Science* 6: 123–128.

Dowe, P. (1995), What's right and what's wrong with transference theories, *Erkenntnis* 42: 363–374.

Dowe, P. (1999), Good connections: Causation and causal processes, in H. Sankey (ed.), *Causation and Laws of Nature,* Dordrecht: Kluwer, pp. 247–63.

Dowe, P. (2004), Causes are physically connected to their effects: Why preventers and omissions are not causes in C. Hitchcock (ed.) *Contemporary Debates in Philosophy of Science*, London Blackwell, pp. 189–96.

Dowe, P. (2007), Causal processes, in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Stanford University. http://plato.stanford.edu/entries/causation-process/causation-process.html; ISSN 1095-5054, revised edition.

Dowe, P. (2009), Absences, possible causation, and the problem of non-locality *Monist* 92: 24–41.

Fair, D. (1979), Causation and the flow of energy, *Erkenntnis* 14: 219–50.

Frayn, K. (2000), Visceral fat and insulin resistance – causative or correlative? *British Journal of Nutrition* 83, Suppl. 1: S71–S77.

Handfield, T., Twardy, C., Korb, K., and Oppy, G. (2008) The metaphysics of causal models: Where's the biff?, *Erkenntnis* 68: 149–168.

Hitchcock, C. (2001), The intransitivity of causation revealed in equations and graphs, *Journal of Philosophy* 98: 273–299.

Hitchcock, C. (2007), Prevention, preemption, and the principle of sufficient reason, *Philosophical Review* 116: 495–532.

Lewis, D. (2004), Causation as influence, in J. Collins, N. Hall and L. Paul, (eds.), *Causation and Counterfactuals,* Cambridge, Mass.: MIT Press, pp. 75–106

Menzies, P. (2004), Difference-making in context, in J. Collins, N. Hall and L. Paul, eds., *Causation and Counterfactuals,* Cambridge, Mass.: MIT Press, pp. 139–80.

Mill, J. S. (1843), *A System of Logic,* London: Longman, Book III Chapter V.

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

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

Salmon, W. (1994), Causality without counterfactuals, *Philosophy of Science* 61: 297–312.

Salmon, W. (1997), Causality and explanation: A reply to two critiques, *Philosophy of Science,* 64: 461–77.

Salmon, W. (1998), *Causality and Explanation,* New York: Oxford University Press.

Spirtes, P., Glymour, C., and Scheines, R. (2000), *Causation, Prediction, and Search. Springer Lecture Notes in Statistics.* 2nd revised edition, Cambridge, MA: MIT Press.

Suppes, P. (1970), *A Probabilistic Theory of Causality,* Amsterdam : North Holland.

Woodward, J. (2003), *Making Things Happen,* New York: Oxford University Press.