11

Difference-making, Closure, and Exclusion

Brad Weslake

1 Introduction

Consider the following two causal exclusion principles:[1]

EXC-SUF If a property *N* is causally sufficient for a property *B*, then no distinct property *M* that supervenes on *N* is a cause of *B*.

EXC-DIF For all distinct properties *M* and *N* such that *M* supervenes on *N*, *M* and *N* do not both cause a property *B*.

I use these labels for properties because of the relevance of these principles to the debate over mental causation: *N* for neural property, *M* for mental property, and *B* for behavioural property. For the remainder of the chapter, I will assume that *M* supervenes on *N*.

What is it for a property to be a cause of, or causally sufficient for, another property? These notions are best introduced by example. Suppose I place two pounds of green pears on the scales, which subsequently reads two pounds (Honderich 1982). On this occasion the property *weighing two pounds* caused, and was causally sufficient for, the property *reading two pounds*. The property *being green* did not cause, and was not causally sufficient for, the property *reading two pounds*.

Opinion divides on how to further analyse these notions. Kim (2005: 35, 39) and List and Menzies (2009: 475, n. 2) treat this sort of talk as elliptical for property instances causing, or being causally sufficient for, other property instances. On this view it is strictly speaking the particular instance of *weighing two pounds* that caused, and was causally sufficient for, the particular instance of *reading two pounds*. More common has been to take causation to be a relation between events, and to

[1] Thanks to audiences at ANU, Hobart, Macquarie, and Sydney, to David Braddon-Mitchell, and especially to Helen Beebee and Peter Menzies for their comments on earlier versions of this paper. Peter's support and generosity have made an enormous impact on my work, and it is an honour to contribute to a volume dedicated to him.

understand this sort of talk as elliptical for properties of the cause being relevant to its causing, or being causally sufficient for, an effect with other properties (Lepore and Loewer 1987; Braun 1995). On this view it is strictly speaking the property *weighing two pounds* that was relevant to the event of the pears being placed on the scales causing, and being causally sufficient for, the event with the property *reading two pounds*. The examples are more compelling than the analyses, and my discussion will not depend on which is correct.

How should we determine whether EXC-SUF and EXC-DIF are true? Many philosophers have supposed that we should do so by reflecting on our concepts of causation and supervenience. Jaegwon Kim (1998), for example, has famously argued that EXC-SUF and EXC-DIF are analytically true. List and Menzies (2009), on the other hand, have argued that EXC-SUF and EXC-DIF are analytically false (see also Menzies and List 2010; in what follows I will refer to these papers together as LM). The explanation for these philosophers reaching contradictory conclusions is that their arguments presuppose different conceptions of causation. Kim assumes a conception of causation as something like production or generation and shows that this notion entails that EXC-SUF and EXC-DIF are true. LM, on the other hand, assume a conception of causation as difference-making, and show that there are possible situations in which EXC-SUF and EXC-DIF are false.

These arguments are illuminating, as they show that different theories of causation generate different implications concerning the possibility of supervenient causation. But we should not lose sight of the fact that there is an alternative way to determine the truth of these principles. Instead of approaching the question analytically, we can see whether there is evidence for or against the principles provided by actual examples. For example, we have good evidence that all actually caused events have causally sufficient physical conditions (Papineau 2001), and also good evidence that some actual events are caused by properties that are distinct from and supervene on those conditions. So we have good evidence that EXC-SUF is false, evidence that does not depend on being able to articulate a theory of causation. Call this *the argument from example against* EXC-SUF. The argument from example against EXC-SUF in turn provides evidence against the notion of causation as production presupposed by Kim, and for the notion of causation as difference-making presupposed by LM.

It appears straightforward to extend this line of argument to EXC-DIF. For the evidence that all actually caused events have causally sufficient physical conditions also appears to be evidence that all actually caused events have physical causes. In other words, to the extent there is evidence for the causal closure of the physical world, it appears to be evidence for closure principles formulated both in terms of causal sufficiency and in terms of causation proper. So the evidence that EXC-SUF is false also appears to be evidence that EXC-DIF is false. Call this *the argument from example against* EXC-DIF. Moreover, extending the argument in this way is warranted by an independently plausible principle:

CAU-SUF If a property *N* is causally sufficient for a property *B*, then *N* is a cause of *B*.

Indeed, it is tempting to regard CAU-SUF as analytic. What could discriminate between, say, nomological and causal sufficiency, if not that causal sufficiency is a variety of sufficiency possessed by causes?[2]

It is therefore surprising that according to LM, the argument from example against EXC-DIF is unsound. LM do agree that EXC-DIF is false. However, they show that it follows from their conception of causation as difference-making that 'the systems that falsify it are very special' (2009: 500), in a way that I will describe below. And as it happens, many of the examples that seem to support the argument from example against EXC-DIF do not fall within this special class. Instead it turns out that, according to LM, they are examples where causally sufficient physical conditions are not causes, and therefore where CAU-SUF is false. In other words, LM may endorse the first of the following causal closure principles, but are committed to rejecting the second:

CLO-SUF Every event for which there exists a causally sufficient property has a causally sufficient physical property.

CLO-DIF Every event that has a cause has a physical cause.

In sum, LM are committed by their difference-making conception of causation to rejecting the argument from example against EXC-DIF, rejecting CAU-SUF, and rejecting CLO-DIF. In what follows, I will argue that we should accept all three. My central claim is that there is an alternative and superior difference-making conception of causation that permits us to accept all three, and hence that we should do so. I begin, in §§2–3, by presenting the LM conception of causation as difference-making. In §2, I describe the account LM give of our causal judgements concerning two examples. In §3 I describe the account LM give of causation as difference-making, and present some of the results concerning exclusion principles that they prove from the account. I then turn, in §§4–6, to criticism of the LM account. In §4 I argue that their notion of difference-making is not well motivated by their own examples, describe a better-motivated notion, and show that both notions are inconsistent with an alternative account of difference-making. I then argue that the alternative account is to be preferred, both on independent grounds and on grounds that it is consistent with the argument from example against EXC-DIF, with CAU-SUF, and with CLO-DIF. In §5, I present an independently plausible conception of causal sufficiency, and argue that it entails that CAU-SUF is true. In §6 I argue that the judgements appealed to by LM are best accounted for pragmatically rather than semantically. I conclude in §7.

[2] Of course, there are notions of causal sufficiency that are more broad than this, for example notions on which effects may be sufficient for causes, or common effects sufficient for each other. But CAU-SUF is, I claim, plausible for the variety of causal sufficiency at play in exclusion arguments.

2 List and Menzies on Causal Judgement

The account of difference-making endorsed by LM is motivated by two examples with a common structure. First is an example introduced by Woodward (2008: 238), idealized from the work of Musallam et al. (2004). Musallam et al. (2004) showed that in macaque monkeys, intentions to reach for particular goals are highly correlated with the aggregate firing rates of neurons in the parietal reach region of the posterior parietal cortex. However, conditional on those aggregate rates, the specific firing rates of the relevant individual neurons are not correlated with those intentions. Suppose that on a particular occasion, Sylvester the macaque monkey reaches for a goal after his parietal reach region neurons fire with particular pattern N_i and aggregate pattern I_i (where N_i entails I_i). Second is an example introduced by Yablo (1992b: 257). Sophie the pigeon pecks at all and only the red things. Conditional on whether a presented object is red, the specific shade of the object is not correlated with pecking. Suppose that on a particular occasion, Sophie pecks at a crimson thing. In what follows, I follow LM in supposing that it is a harmless idealization to model these cases deterministically, in the sense that Sylvester reaches when and only when his neurons fire with pattern I_i and Sophie pecks when and only when the object is red.

Now consider the following candidate explanations:

SYLV_1 Sylvester reached because his neurons fired with pattern N_i.
SYLV_2 Sylvester reached because his neurons fired with pattern I_i.
SOPH_1 Sophie pecked because the object was crimson.
SOPH_2 Sophie pecked because the object was red.

A natural reaction to these examples is that there is a respect in which SYLV_2 provides a better explanation that SYLV_1, and in which SOPH_2 provides a better explanation than SOPH_1. Moreover, a natural hypothesis for what makes this explanatory difference is that in each case, there exist alternatives to the lower-level property, consistent with the higher-level property, that would have led to the same effect. That is, Sylvester would have reached had (for example) his neurons fired with pattern N_j (where N_j entails I_i), and Sophie would have pecked had (for example) she been presented with a scarlet object. I will argue in §6 that this reaction and hypothesis are both correct.

Do these differences in explanatory status reflect a difference in causal status? According to LM, following Yablo (1992a, 1992b, 1997, 2003, 2005), they do. Consider the following propositions concerning causation:

SYLV_3 Sylvester's neurons firing with pattern N_i caused his reaching.
SYLV_4 Sylvester's neurons firing with pattern I_i caused his reaching.
SOPH_3 The object being crimson caused Sophie's pecking.
SOPH_4 The object being red caused Sophie's pecking.

And consider the following propositions concerning difference-making:

SYLV$_5$ Sylvester's neurons firing with pattern N_i made a difference to his reaching.
SYLV$_6$ Sylvester's neurons firing with pattern I_i made a difference to his reaching.
SOPH$_5$ The object being crimson made a difference to Sophie's pecking.
SOPH$_6$ The object being red made a difference to Sophie's pecking.

We are more inclined to assert SYLV$_4$ than SYLV$_3$, and SOPH$_4$ than SOPH$_3$. According to LM, this is because SYLV$_4$ and SOPH$_4$ are both true, while SYLV$_3$ and SOPH$_3$ are both false. In a term introduced by Yablo (1992a), their claim is that causes must be *proportional* to their effects. This in turn is supposed to be explained by the fact that there is a notion of difference-making relevant to analysing causation on which SYLV$_6$ and SOPH$_6$ are both true, while SYLV$_5$ and SOPH$_5$ are both false.

3 List and Menzies on Difference-making

LM propose an account of difference-making that is intended to make sense of the causal and explanatory judgements prompted by examples such as those involving Sylvester and Sophie. They then prove that the account entails a number of very interesting results concerning EXC-DIF. In this section I present the account and the results.

LM present two formulations of the intuition that causes should make a difference to their effects. The two formulations, which they claim are equivalent, are as follows:

TRUE-DIFF$_1$ The presence of F makes a difference to the presence of G in the actual situation just in case (i) if any relevantly similar possible situation instantiates F, it instantiates G; and (ii) if any relevantly similar possible situation instantiates $\neg F$, it instantiates $\neg G$.

TRUE-DIFF$_2$ The presence of F makes a difference to the presence of G in the actual world if and only if it is true in the actual world that (i) F is present $\Box\!\!\rightarrow G$ is present; and (ii) F is absent $\Box\!\!\rightarrow G$ is absent.

In the following section, I will argue that these formulations are not equivalent, and suggest a different way of formulating the idea behind TRUE-DIFF$_1$. In this section I focus on TRUE-DIFF$_2$. LM prove a number of very interesting results concerning TRUE-DIFF$_2$. Their proofs depend on a possible-worlds semantics for counterfactuals that is similar but not identical to the semantics developed by Lewis (1973), and I recommend their paper to readers who are interested in the details. Here I wish to highlight the most important results. The first concerns situations in which EXC-DIF is false:

Compatibility result M and N both make a difference to B iff (i) B is present in all closest M-worlds; (ii) B is absent in all closest $\neg M$-worlds; and (iii) B is absent in all closest $\neg N$-worlds that are M-worlds.

It is this result that shows that EXC-DIF is not true in general, if causation is analysed in terms of TRUE-DIFF$_2$. Call a causal relation between M and B *realization-sensitive* iff in

all those M-worlds that are closest $\neg N$-worlds, B is no longer present. Then this result entails that EXC-DIF is false whenever some higher-level property stands in a realization-sensitive causal relation to another property.

Here is an example to illustrate these concepts. Suppose that my intention to dial a certain phone number is realized by a certain neural pattern. Suppose moreover that my intention is a difference-maker, in the sense of TRUE-DIFF$_2$, for two effects: my dialling the number, and my hand shaking. This means that in all the closest worlds in which the intention is present, both the dialling and the shaking are present; and that in all the closest worlds worlds in which the intention is absent, both the dialling and the shaking are absent. There may nevertheless be a difference in how realization-sensitive these difference-making relations are to the presence or absence of the neural pattern. Suppose that in the nearest worlds in which the intention is present but the neural pattern is absent, the dialling occurs but the shaking doesn't. Then the relation of the intention to the dialling is realization-insensitive and therefore excludes the neural pattern from making a difference to the dialling; while the relation of the intention to the shaking is realization-sensitive and therefore does not exclude the neural pattern from making a difference to the shaking.[3]

The following is an immediate corollary of *Compatibility result*:

> *Incompatibility result* EXC-DIF holds if and only if either (i) B is absent in some closest M-worlds, or (ii) B is present in some closest $\neg M$-worlds, or (iii) B is present in some closest $\neg N$-worlds that are M-worlds.

If conditions (i) or (ii) are met then M is excluded (*upwards exclusion*) while if (iii) is met then N is excluded (*downwards exclusion*).

The example of the intention and the dialling exhibited a case of downwards exclusion. Here is an example of upwards exclusion. Suppose that my feeling nervous is realized by a certain neural pattern. Suppose moreover that the neural pattern is a difference-maker, in the sense of TRUE-DIFF$_2$, for the value of a certain electromagnetic field. If in some of the closest worlds where I feel nervous, the value of the electromagnetic field is different, then my feeling nervous is excluded as a difference-maker for the value of the electromagnetic field.

These are very important results. For as LM argue, it is plausible that in science we frequently seek realization-insensitive causal relations. If so, and if they are right that TRUE-DIFF$_2$ is relevant to the analysis of causation, then many causal relations discovered in science exclude their realizers—including those motivating the argument from example against EXC-DIF. According to that argument, recall, we have empirical

[3] The example shows that a single property may bear a realization-insensitive difference-making relation to one property and a realization-sensitive difference-making relation to another property. This suggests that the LM thought that realization-sensitivity might be 'a plausible criterion for identifying higher-level properties with their physical realizers' (2009: 477 n. 5) would need to be qualified in order to be plausible.

evidence for the claim that all actually caused events have physical causes, and also for the claim that some actual events are caused by properties that are distinct from and supervene on those causes. In response to this argument, LM must say that our evidence for actual events being caused by supervening properties consists mainly of evidence for realization-insensitive relations between those properties. Since these exclude their subvening properties from causing the same effects, we must deny the claim that all actually caused events have physical causes, and hence deny a crucial premise in the argument from example against EXC-DIF. It should now be clear, then, why LM say that the systems that falsify EXC-DIF are special: they are special in the sense that they involve the sorts of realization-sensitive causal relations on which science does not focus.

4 An Alternative Conception of Difference-making

As I noted above, a natural hypothesis for the basis of our explanatory judgements in cases like Sylvester and Sophie is that there are alternatives to the lower-level properties that would have led to the same effect. TRUE-DIFF$_1$ provides a natural formulation of this hypothesis as a principle concerning difference-making, requiring that in relevantly similar situations, the cause covaries with the effect. Notice however that TRUE-DIFF$_2$ is different, instead requiring the truth of a counterfactual concerning what would have happened had the candidate difference-making property been absent. Now there is something strange about the appeal to such counterfactuals in the cases that motivate the appeal to proportionality.[4] Take the case of Sophie. According to TRUE-DIFF$_2$, whether or not crimson is a difference-maker depends on whether the nearest worlds in which the patch is not crimson are ones in which it is still red. But this cannot be responsible for our causal or explanatory judgements, for in the description of the example offered by both Yablo and LM, it is underdetermined whether this counterfactual is true or false. Regarding the example, LM write: 'It is natural to interpret these counterfactuals in terms of a similarity relation that makes the closest worlds in which the target is not crimson ones where it is some other shade of red' (2009: 488). But this is not natural at all, for in the story as told by both Yablo and LM, we are told nothing about the way in which the patch came to be crimson.[5]

LM might reply that even though the truth of the relevant counterfactual is underdetermined by the description of the example, we naturally assume that it is true nonetheless. But this cannot be right, for I claim that our causal and explanatory judgements are robust across variations of the example in which the truth value of the counterfactual is explicitly varied. Consider the following two variations of the

[4] I develop the same line of criticism against Yablo in Weslake (2013).

[5] This is also noted by Shapiro (2012), though he seems to think that it is natural in the neural rates cases though unnatural in other cases.

example. In the first, we stipulate that had the patch not been crimson it would have been some other shade of red. In the second, we stipulate that had the patch not been crimson it would not have been some other shade of red. I claim that our causal and explanatory judgements are identical across these examples. If so, then the LM formulation of difference-making in terms of these counterfactuals is mistaken. I conclude that TRUE-DIFF$_2$ is inequivalent and inferior to TRUE-DIFF$_1$, as a conception of difference-making responsible for our causal judgements.[6]

Here is an alternative formulation of difference-making that is immune from this line of argument, and is therefore better suited to play the role in our causal judgements demanded by LM:

> TRUE-DIFF$_3$ The presence of *F* makes a difference to the presence of *G* in the actual situation just in case (i) for all relevant ways *F* could have been instantiated, *G* would have been instantiated; and (ii) for all relevant ways ¬*F* could have been instantiated, ¬*G* would have been instantiated.

Since it does not appeal to counterfactuals concerning what would have happened had the candidate difference-making property been different, but rather to counterfactuals concerning what would have happened had relevant alternatives obtained, TRUE-DIFF$_3$ is more plausibly equivalent to TRUE-DIFF$_1$.[7] However as I will now argue, the problem with TRUE-DIFF$_3$ is that it is inconsistent with an alternative, and more plausible, account of difference-making.

LM introduce their conception of difference-making by suggesting that it is compatible with a range of different theories of causation: 'Since a conception of this kind is common to several different theories of causation—for example, counterfactual, interventionist, and contrastive ones—our use of it in investigating the exclusion principle should be congenial to a broad range of such theories' (476). They also restrict their discussion in a number of ways, claiming that it does not affect the generality of their conclusions. Most important for present purposes is the following: 'we discuss causal relations involving properties. Causation is best understood, we believe, as a relation between variables. So causation involving properties is a special case in which the variables are binary. A more general treatment would handle causation involving many-valued variables' (478). However, it turns out that the

[6] LM might reply that the similarity metric governing our causal judgements is not the same as the similarity metric responsible for our ordinary counterfactual judgements. In that case, since TRUE-DIFF$_1$ is supposed to be equivalent to TRUE-DIFF$_2$, my criticism of TRUE-DIFF$_3$ below also amounts to a criticism of TRUE-DIFF$_2$.

[7] The restriction to relevant ways in which properties could have been instantiated is required in order for TRUE-DIFF$_3$ to be a candidate for equivalence with TRUE-DIFF$_1$. There are of course possibilities in which, for example, the patch is red but Sophie doesn't peck because some improbable event disrupts the structure of the situation. The appeal to relevance is supposed to exclude these possibilities from influencing the truth values of the associated counterfactuals. No such restriction is made in the case of TRUE-DIFF$_2$, since here the idea is that the similarity relation itself excludes these possibilities.

more general treatment suggested by their account is importantly different from an alternative account of difference-making.

This is most easily seen by situating the difference-making principles within the interventionist theory of causation, with which I assume readers are familiar.[8] Suppose we make the natural assumption that in an appropriate causal model, all variable values in the model represent relevant alternative possibilities. Then TRUE-DIFF$_3$ can be reformulated as follows:

> TRUE-DIFF$_4$ Variable value $F = f$ makes a difference to variable value $G = g$ in the actual situation just in case (i) an intervention setting $F = f$ would result in $G = g$; and (ii) for all variable values $F = f'$ (where $f \neq f'$), an intervention setting $F = f'$ would result in $G = g'$ (where $g \neq g'$).

This contrasts with the following difference-making principle:

> TRUE-DIFF$_5$ Variable value $F = f$ makes a difference to variable value $G = g$ in the actual situation just in case (i) an intervention setting $F = f$ would result in $G = g$; and (ii) for some variable value $F = f'$ (where $f \neq f'$), an intervention setting $F = f'$ would result in $G = g'$ (where $g \neq g'$).

While TRUE-DIFF$_4$ requires that all interventions would make a difference, TRUE-DIFF$_5$ requires merely that there exist an intervention that would make a difference.[9] This difference between TRUE-DIFF$_4$ and TRUE-DIFF$_5$ is obscured by a focus on causal models with binary variable values, where they collapse.

Now it is principles such as TRUE-DIFF$_5$ that have played a role in all theories of causation that have been proposed in the interventionist literature.[10] The interventionist theories of causation proposed by Hitchcock (2001), Woodward (2003: §2.7), and Halpern and Pearl (2005) all require merely that there exist an intervention that makes a difference, not that all interventions make a difference. According to all of these theories, if there is an appropriate causal model in which there is an alternative to crimson that would not have led to Sophie's pecking, then crimson caused Sophie to peck. According to TRUE-DIFF$_4$ however, since there is an appropriate causal model in which there is an alternative to crimson that would also have led to Sophie's pecking, crimson did not make a difference to Sophie's pecking.

In the remainder of this section I present three reasons to prefer TRUE-DIFF$_5$ to TRUE-DIFF$_4$, as a conception of difference-making relevant to the analysis of causation.

First, as noted by both Woodward (2003: 66) and Shapiro and Sober (2012), TRUE-DIFF$_5$ delivers better verdicts when the relationship of counterfactual dependence

[8] For a comprehensive philosophical overview see Woodward (2003), for a more technical presentation see Pearl (2009), and for brief introductions see Hitchcock (2001, 2007) or Weslake (forthcoming).

[9] The difference between these principles is also noted by Marras and Yli-Vakkuri (2010).

[10] I do not claim to have made a complete survey. Both here and below, where I make a similar claim, I support my case by discussing the three most prominent theories in the literature. In Weslake (forthcoming), I show how these theories relate to one another and argue for an alternative.

between variables maps multiple values of one variable onto a single value of another variable. Consider for example the relationship between the amount of water given to a plant during a particular period, and whether the plant lives or dies.[11] The plant will die if no water is given, survive if a volume of water within some particular range is given, and die if a volume of water above that range is given. Suppose on some particular occasion, no water is given. One natural way to model this situation is with a three-valued variable representing the volume of water given ($W = 0$ when no water, $W = 1$ when within range, $W = 2$ when above range), and a two-valued variable representing whether the plant lives or dies ($P = 0$ when the plant dies, $P = 1$ when the plant survives). In the actual case, $W = 0$ and $P = 0$. According to TRUE-DIFF$_5$, not watering the plant made a difference to the plant's dying, since had the plant been watered within range, it would have survived. But according to TRUE-DIFF$_4$, not watering the plant did not make a difference to the plant's dying, since had it been watered above range, it would still have died. An adherent of TRUE-DIFF$_4$ faces a choice over how to represent the difference-maker in this situation. They might say that models which map multiple values of one variable onto a single value of another are for that reason inappropriate. In this case, the correct model would require a binary variable, one value of which represents the disjunctive property of the plant's either being not watered or watered above range. Alternatively, they might say that while the three-valued variable may appear in the model, the disjunctive difference-maker for the plant's dying is represented by the disjunction of variable values $W = (0 \vee 2)$. Either way, we have a problem. For setting aside worries about the causal status of omissions, it is clear that not watering the plant caused it to die. It is equally clear that the causal status of the disjunctive property is at best questionable. So TRUE-DIFF$_5$ is preferable to TRUE-DIFF$_4$.[12] Second, TRUE-DIFF$_5$ is better suited to a contrastive theory of causation than TRUE-DIFF$_4$. One of the central attractions of a contrastive theory is the way it allows us to say, in the plant case, that not watering the plant rather than watering within range was a cause of death, while not watering the plant rather than watering above range was not a cause of death.[13] TRUE-DIFF$_5$ can be adapted in the obvious and natural way to capture the corresponding claims about difference-making. It is difficult to see how to do the same for TRUE-DIFF$_4$.[14]

[11] A similar example is used to make a different point by Woodward and Hitchcock (2003: 183).

[12] This argument depends on the assumption that there exists a model that represents all three possibilities for watering the plant and that is capable of representing the cause of death. An adherent of TRUE-DIFF$_4$ could reject this assumption, and argue that TRUE-DIFF$_4$ delivers the right result with respect to a model that does not represent the possibility of the plant being watered above range. However, it is hard to see how this move could be independently motivated. Thanks here to Helen Beebee.

[13] See for example Hitchcock (1996: §II) and Maslen (2004).

[14] Contrastive claims of this sort are also hard to square with the claim made by LM that SYLV$_3$ and SOPH$_3$ are false. For example, it seems correct to say that the object being crimson rather than black caused Sophie to peck. How then can it also be correct to say that the object being crimson did not cause Sophie to peck? I return to the question of whether SYLV$_3$ and SOPH$_3$ are false in §6.

Third, TRUE-DIFF$_5$ is consistent with CAU-SUF, with CLO-DIF, and therefore with the argument from example against EXC-DIF. To the extent that CAU-SUF, CLO-DIF, and the argument from example against EXC-DIF are independently plausible, we therefore also have further grounds for preferring TRUE-DIFF$_5$ to TRUE-DIFF$_4$. In the following section I argue that TRUE-DIFF$_5$ entails CAU-SUF. But without going through that argument, the consistency of TRUE-DIFF$_5$ with these principles is easy to see. The reason that TRUE-DIFF$_4$ required giving them up was that some of the closest worlds in which some actually sufficient physical condition doesn't occur may be worlds in which the candidate effect occurs nonetheless. According to TRUE-DIFF$_4$, this undermines the status of the physical condition as a difference-maker. But according to TRUE-DIFF$_5$ it does not. Rather, all that is required is that there be some relevant alternative to the physical condition that would not have led to the same effect.

5 Two Definitions of Causal Sufficiency

In this section I argue that CAU-SUF is a consequence of an independently plausible conception of causal sufficiency that can be formulated using the interventionist framework. I will make use of the following definitions and assumptions. I will refer to a possible assignment of values to all variables in a model as a *state* of the model. I will refer to variables that have no parents as *exogenous*, and variables that have parents as *endogenous*. I will refer to a model that has the variables in model M as a subset an *expansion* of M. Finally, I will assume that the equations specifying the relations of counterfactual dependence between variables are all deterministic. I first present the conception of causal sufficiency, and then present an argument that it entails CAU-SUF.

First we need the concept of a redundancy range (Woodward 2003: 83):

> REDUNDANCY For variable values $X = x$ and $Y = y$ in model M, define $V_1 \ldots V_n$ as all other variables in M. Values $V' \ldots v_n$ are on the *redundancy range* for $V_1 \ldots V_n$ with respect to $X = x$ and $Y = y$ iff no intervention setting $V_1 \ldots V_n$ to $v_1 \ldots v_n$ while holding fixed $X = x$ would result in $Y = y'$, where $y \neq y'$.

A natural conception of causal sufficiency can then be defined as follows:

> SUFFICIENCY $X = x$ is *causally sufficient* for $Y = y$ in model M iff (i) Y is an endogenous variable; and (ii) all possible combinations of values $v_1 \ldots v_n$ for all other variables $V_1 \ldots V_n$ in M are on the redundancy range with respect to $X = x$ and $Y = y$.

Less formally, the definition says that a first variable is causally sufficient for a second just in case, no matter what values other variables in the model had taken, the second would have taken the value it did.

This definition is limited in two ways. First, it is a notion of causal sufficiency defined for values of single variables. There is a natural generalization to values of

multiple variables, but I leave this for another occasion. Second, it is a consequence of the definition that only the values of immediate parent variables will be causally sufficient for a given variable value. Again, there is a natural generalization to permit non-immediate parent variable values to be causally sufficient, but I also leave this for another occasion. Finally, note that this is a model-relative notion of causal sufficiency. According to the definition, a variable may be causally sufficient for another relative to one model, but not to another model. A more strict, though still model-relative notion can be defined as follows:

ÜBER-SUFFICIENCY $X = x$ is *über-sufficient* for $Y = y$ in model M iff $X = x$ is causally sufficient for $Y = y$ in all expansions of M.

Less formally, the definition says that a first variable is über-sufficient for a second just in case the first is causally sufficient for the second both in the model in question, and in all other models containing the variables in the model in question as a subset.

The model-relativity of the first definition is attractive. Suppose for instance that there is a neural property that is required to be instantiated in order for Sophie to peck. Relative to a model that does not include a variable representing this property, crimson may be causally sufficient for pecking. Relative to a model that does include such a variable, crimson will not be causally sufficient for pecking. This flexibility is helpful in making sense of the ways that our judgements of causal sufficiency can be context-sensitive. On this account, this context-sensitivity traces to differences in which alternative possibilities are relevant in a given context. For example, whether crimson is judged to be causally sufficient depends on whether the neural property in question is treated as something fixed or instead as something capable of variation.

The notion of causal sufficiency is therefore useful for capturing the idea of sufficiency in the circumstances, where what counts as the circumstances may shift with context. The notion of über-sufficiency, in turn, is useful for capturing a notion of sufficiency strictly speaking. For example, crimson is not strictly speaking causally sufficient for pecking, for to speak strictly is to treat everything as capable of variation.[15]

I will now argue that by the lights of all theories of causation that have been proposed in the interventionist literature, if $X = x$ is causally sufficient for $Y = y$ in the defined sense, then $X = x$ is a cause of $Y = y$. It follows from the definition of causal sufficiency that $Y = y$ is an endogenous variable. And as noted above, it is also a consequence of the definition that X is an immediate parent of Y. It follows from Y being endogenous and the equations being deterministic that there is an alternative state of the model in which the immediate parents of Y have values different from their actual values, and in which $Y = y'$, where $y \neq y'$. It follows from $X = x$ being causally sufficient for $Y = y$ that this alternative state must be one in which $X = x'$, where $x \neq x'$. Finally, it follows from the definitions of causation proposed by

[15] I am grateful to Helen Beebee for suggesting the connection with the idea of sufficiency in the circumstances.

Hitchcock (2001), Woodward (2003: §2.7), and Halpern and Pearl (2005) that in virtue of the existence of this alternative state, $X = x$ is a cause of $Y = y$.[16]

I draw two conclusions from this result. First, we have an independent argument for CAU-SUF, and therefore an independent reason to reject the LM conception of difference-making. Recall that according to CAU-SUF, causally sufficient properties are causes. If so, then LM cannot consistently accept a closure principle formulated in terms of causal sufficiency while rejecting a closure principle formulated in terms of causation proper. Second, true-diff$_5$ is consistent with CAU-SUF.

6 Causation and Explanation

I have argued that we should reject the conception of difference-making defended by LM, insofar as it is relevant to the analysis of causation, and instead endorse the conception of difference-making that underlies the theories of causation defended in the interventionist literature. That conception, which can be formulated as TRUE-DIFF$_5$, has the advantage of consistency with CAU-SUF and CLO-DIF. It also undermines the LM arguments for both upwards and downwards exclusion, since those arguments depend on assuming TRUE-DIFF$_2$.

This leaves open the question of what to say about the judgements that motivated LM, introduced in §2. For LM are correct that we are more inclined to assert SYLV$_4$ than SYLV$_3$, and SOPH$_4$ than SOPH$_3$. According to LM, this is because SYLV$_4$ and SOPH$_4$ are both true, while SYLV$_3$ and SOPH$_3$ are both false. But according to interventionist theories of causation, this cannot be right. For there are relevant alternatives to crimson that would have led to Sophie not pecking, and relevant alternatives to N_i that would have led to Sylvester not reaching. According to interventionist theories, the truth of these counterfactuals is sufficient for the truth of SYLV$_3$ and SOPH$_3$.

This rules out the possibility of explaining our judgements semantically. In the remainder of this section, I argue that they should instead be explained pragmatically.[17] In particular, I suggest that there is pragmatic pressure to assert only the most explanatory proposition of the pairs SYLV$_4$ and SYLV$_3$, and SOPH$_4$ and SOPH$_3$, respectively. I will not defend the claim that there is pragmatic pressure to assert only one of each pair, deferring to Swanson (2010). Instead, I will focus on defending the claim that there are important explanatory differences between SYLV$_4$ and SYLV$_3$, and SOPH$_4$ and SOPH$_3$, respectively.

I will focus on the case of Sophie, and argue that there are three important respects in which SOPH$_4$ provides a better explanation than SOPH$_3$. I first present the dimensions of

[16] I will not work through the details here, but the easiest way to see this is to consult the formulations of these theories in Weslake (forthcoming), and to keep in mind that the differences between them only arise when a candidate cause is not an immediate parent of a candidate effect.

[17] In what follows, I draw on arguments first presented in Weslake (2013).

explanation, and then argue that SOPH$_4$ provides a better explanation than SOPH$_3$ along each dimension.

A first dimension of explanatory value is identified by Woodward and Hitchcock (2003), who argue that an explanation is better to the extent it specifies more answers to questions concerning how the explanandum would have been different, had the explanans been different. I will call this dimension of explanatory value *dependency*. A second dimension of explanatory value is identified in Weslake (2010), where I argue that an explanation is better to the extent that it would apply to a wider range of possible situations. I will call this dimension of explanatory value *abstraction*. A third dimension of explanatory value, which can be extracted from Woodward (2006), is that an explanation is better to the extent it would continue to obtain under various changes to the actual circumstances. I will call this dimension of explanatory value *insensitivity*.[18]

Since on interventionist theories of causation it is TRUE-DIFF$_5$ that grounds the truth of causal claims, to assert merely that one variable value is a cause of another is not to be especially informative. For it is to assert merely that there exists an alternative to that value which would have led to a difference to the effect.[19] It is much more informative to learn the exact nature of the dependency relation between the two variables: to learn the complete mapping from alternative values of the cause to alternative values of the effect (and more generally, to learn if and how this in turn depends on the values of other variables). Since dependency is a dimension of explanatory value, to possess this information is thereby to possess a better explanation. Now one way to possess information of this sort is to know that TRUE-DIFF$_4$ is true of a cause. So on the assumption that when asserting causal propositions one should be maximally informative with respect to factors that make for explanatory differences, if there is a choice between citing two causes, one of which satisfies TRUE-DIFF$_4$ and one of which does not, one should cite the cause that does. One should therefore assert SOPH$_4$ rather than SOPH$_3$. This is not because SOPH$_4$ is true and SOPH$_3$ is false, but because asserting a causal proposition conversationally implicates that the cause satisfies TRUE-DIFF$_4$. And this in turn is not because TRUE-DIFF$_4$ plays a role in the analysis of causation, but rather because it plays a role in the analysis of explanation.[20]

It is straightforward to see that SOPH$_4$ is superior to SOPH$_5$ along the dimensions of abstraction and insensitivity. Any situation in which SOPH$_5$ applies is also a situation in which SOPH$_4$ applies, but not vice versa. Hence SOPH$_4$ is better along the dimension of abstraction. Likewise, there are a range of changes to the actual circumstances

[18] Woodward himself does not explicitly make the connection between insensitivity and explanatory value. Instead, he argues directly for the relevance of insensitivity to our causal judgements.

[19] This is an oversimplification, of course. Interventionist theories of causation agree that this is sufficient for causation, but differ concerning the exact difference-making conditions that are necessary for causation.

[20] For a similar diagnosis of the role of proportionality, see Bontly (2005).

under which SOPH$_4$ would continue to obtain but in which SOPH$_5$ would not. For example, had Sophie been presented with a scarlet patch rather than a crimson patch, SOPH$_4$ would still have been true but SOPH$_5$ would have been false. So on the assumption that only the most explanatory proposition of the pair SOPH$_4$ and SOPH$_5$ should be asserted, both abstraction and insensitivity demand that it is SOPH$_4$ rather than SOPH$_5$.

I conclude that there are three independent dimensions of explanatory value that speak in favour of asserting SOPH$_4$ rather than SOPH$_5$. We can thereby explain the judgements that motivate the LM account of difference-making without making that account part of the analysis of causation.[21]

7 Conclusion

I have argued that the LM conception of causation as difference-making should be rejected. It is not well motivated by their favoured examples, it leads to counter-intuitive results in mundane cases of causation appropriately modelled by variables with multiple values, and it is difficult to square with a contrastive theory of causation. Moreover, endorsing it would require giving up the argument from example against EXC-DIF, and rejecting plausible principles concerning causal sufficiency (CAU-SUF) and causal closure (CLO-DIF). It is a great virtue of the work done by LM on causation to have made these consequences clear. I have argued too that there is an alternative conception of causation as difference-making that does not have the same problems. This is the conception of difference-making at the heart of all interventionist theories of causation. This conception better handles variables with multiple values, better fits with a contrastive theory of causation, and is consistent with the arguments and principles that the LM conception would require us to reject. Indeed, I argued that CAU-SUF is entailed by the interventionist theory in conjunction with an independently plausible account of causal sufficiency. Finally, I argued that our judgements concerning the examples used to motivate the LM conception should be explained pragmatically rather than semantically. This does not mean that the difference-making principles proposed by LM play no role in our judgements, but rather that they are best seen as part of the theory of explanatory value rather than principles at the heart of the analysis of causation.

References

Bontly, T. D. 2005. 'Proportionality, Causation, and Exclusion', *Philosophia*, 32: 331–48.
Braun, D. 1995. 'Causally Relevant Properties', *Philosophical Perspectives*, 9: 447–75.

[21] It is also worth noting that while we are disinclined to assert SOPH$_5$, we are not inclined to assert that it is false (Maslen forthcoming).

Halpern, J. Y., and Pearl, J. 2001. 'Causes and Explanations: A Structural-Model Approach. Part I: Causes', in *Uncertainty in Artificial Intelligence: Proceedings of the Seventeenth Conference*, ed. J. Breese and D. Koller. San Francisco: Morgan Kaufmann, 194–202. Revised version published as Halpern and Pearl (2005).

Halpern, J. Y., and Pearl, J. 2005. 'Causes and Explanations: A Structural-Model Approach. Part I: Causes', *British Journal for the Philosophy of Science*, 56: 843–87. Revised version of Halpern and Pearl (2001).

Hitchcock, C. 1996. 'Farewell to Binary Causation', *Canadian Journal of Philosophy*, 26: 267–82.

Hitchcock, C. 2001. 'The Intransitivity of Causation Revealed in Equations and Graphs', *The Journal of Philosophy*, 98: 273–99.

Hitchcock, C. 2007. 'Prevention, Preemption, and the Principle of Sufficient Reason', *Philosophical Review*, 116: 495–532.

Honderich, T. 1982. 'The Argument for Anomalous Monism', *Analysis*, 42: 59–64.

Kim, J. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim, J. 2005. *Physicalism, Or Something Near Enough*. Princeton: Princeton University Press.

Lepore, E., and Loewer, B. 1987. 'Mind Matters', *The Journal of Philosophy*, 84: 630–42.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.

List, C., and Menzies, P. 2009. 'Nonreductive Physicalism and the Limits of the Exclusion Principle', *Journal of Philosophy*, 106: 475–502.

Marras, A., and Yli-Vakkuri, J. 2010. 'Causal and Explanatory Autonomy: Comments on Menzies and List', in *Emergence in Mind*, ed. C. Macdonald and G. Macdonald. Oxford: Oxford University Press, 129–38.

Maslen, C. 2004. 'Causes, Contrasts, and the Nontransitivity of Causation', in *Counterfactuals and Causation*, ed. J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: MIT Press, 341–57.

Maslen, C. Forthcoming. 'Pragmatic Explanations of the Proportionality Constraint on Causation', in *Making a Difference*, ed. H. Beebee, C. Hitchcock, and H. Price. Oxford: Oxford University Press.

Menzies, P., and List, C. 2010. 'The Causal Autonomy of the Special Sciences', in *Emergence in Mind*, ed. C. Macdonald and G. Macdonald. Oxford: Oxford University Press, 108–28.

Musallam, S., et al. 2004. 'Cognitive Control Signals for Neural Prosthetics', *Science*, 305: 258–62.

Papineau, D. 2001. 'The Rise of Physicalism', in *Physicalism and its Discontents*, ed. C. Gillett and B. Loewer. Cambridge: Cambridge University Press, 3–36. Reprinted in Stone and Wolff (2000: 174–208).

Pearl, J. 2009. *Causality*, Second Edition. Cambridge: Cambridge University Press.

Shapiro, L. A. 2012. 'Mental Manipulations and the Problem of Causal Exclusion', *Australasian Journal of Philosophy*, 90: 507–24.

Shapiro, L. A., and Sober, E. 2012. 'Against Proportionality', *Analysis*, 72: 89–93.

Stone, M. W. F., and Wolff, J. (eds). 2000. *The Proper Ambition of Science*. Vol. 2. London: Routledge.

Swanson, E. 2010. 'Lessons from the Context Sensitivity of Causal Talk', *Journal of Philosophy*, 107: 221–42.

Weslake, B. 2010. 'Explanatory Depth', *Philosophy of Science*, 77: 273–94.

Weslake, B. 2013. 'Proportionality, Contrast and Explanation', *Australasian Journal of Philosophy*, 91: 785–97.

Weslake, B. Forthcoming. 'A Partial Theory of Actual Causation', *British Journal for the Philosophy of Science*.

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

Woodward, J. 2006. 'Sensitive and Insensitive Causation', *The Philosophical Review*, 115: 1–50.

Woodward, J. 2008. 'Mental Causation and Neural Mechanisms', in *Being Reduced: New Essays on Reduction, Explanation and Causation*, ed. J. Kallestrup and J. Hohwy. Oxford: Oxford University Press, 218–62.

Woodward, J., and Hitchcock, C. 2003. 'Explanatory Generalizations, Part II: Plumbing Explanatory Depth', *Noûs*, 37: 181–99.

Yablo, S. 1992a. 'Cause and Essence', *Synthese*, 93: 403–49. Reprinted in Yablo (2010: 59–97).

Yablo, S. 1992b. 'Mental Causation', *The Philosophical Review*, 101: 245–80. Reprinted in Yablo (2009: 222–48).

Yablo, S. 1997. 'Wide Causation', *Noûs*, 31: 251–81. Reprinted in Yablo (2009: 275–306).

Yablo, S. 2003. 'Causal Relevance', *Philosophical Issues*, 13: 316–28. Reprinted in Yablo (2009: 307–19).

Yablo, S. 2005. 'Prime Causation', *Philosophy and Phenomenological Research*, 70: 459–67.

Yablo, S. 2009. *Thoughts: Papers on Mind, Meaning, and Modality*. Oxford: Oxford University Press.

Yablo, S. 2010. *Things: Papers on Objects, Events, and Properties*. Oxford: Oxford University Press.