

Causes as Difference-Makers

Author(s): Carolina Sartorio

Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Mar., 2005, Vol. 123, No. 1/2, Selected Papers from the 2003 Bellingham Conference (Mar., 2005), pp. 71-96

Published by: Springer

Stable URL: <https://www.jstor.org/stable/4321573>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*

CAROLINA SARTORIO

CAUSES AS DIFFERENCE-MAKERS★

1. INTRODUCTION

David Lewis wrote in “Causation:”

We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it.¹

Call this idea, according to which a cause is a “difference-maker,” *the difference-making idea*. The difference-making idea famously motivated Lewis’s counterfactual theory, an attempt to analyze the concept of cause in terms of the relation of counterfactual dependence between events.²

However, as we will see, Lewis’s counterfactual theory ends up misrepresenting the difference-making idea: it counts as causes things that aren’t difference-makers. We should then look for an alternative way of spelling out the difference-making idea. This is what I do in this paper. I make a new proposal on how causes are difference-makers, and I argue that the new proposal succeeds in capturing the difference-making idea.

Two words of clarification are in order. First, the view that I defend here is not an *analysis* of causation. It sets a constraint on the concept of cause, and thus it helps to carve up the concept, while at the same time leaving some room for different ways of pinning it down. Second, this paper is an attempt to establish how *best* to capture the difference-making idea; it is not – at least, not primarily – a defense of the claim that we

★ Special thanks to Juan Comesaña, Ned Hall, Judith Thomson, and Stephen Yablo. Thanks also to Tyler Doggett, Andy Egan, Elizabeth Harman, Daniel Hausman, James John, Sarah McGrath, Agustín Rayo, Gabriel Uzquiano, the participants at the 2003 Bellingham Summer Philosophy Conference, especially my commentator, Douglas Ehring, and two referees for *Philosophical Studies*.

should endorse it. However, at the end of the paper I point to an important advantage of endorsing the difference-making idea: I argue that a concept of cause that results from embracing it is particularly well suited for the work causation does in moral theory.

2. TWO WAYS OF MAKING A DIFFERENCE

I will start by illustrating how the difference-making idea motivated Lewis's counterfactual theory and how, despite this fact, Lewis's theory failed to capture it. This discussion will then serve to motivate my own proposal.

On its first pass, Lewis's counterfactual theory (CT) states that a cause is something without which the effect wouldn't have occurred:

CT (First Pass): C causes E if and only if E counterfactually depends on C, i.e., if C hadn't happened, then E wouldn't have happened.³

That is, a cause makes a difference to its effect in that the effect wouldn't have occurred without the cause. This first pass has obvious counterexamples, such as the following:

Assassination: Assassin shoots Victim and, as a result, Victim dies. However, Backup is waiting in reserve. Had Assassin not shot, Backup would have, and Victim would still have died (in a very similar way, at around the same time, etc.).

Victim's death (intuitively, the *same* death) would still have occurred if Assassin hadn't shot; hence, CT (First Pass) entails that Assassin's shooting wasn't a cause of Victim's death. But, clearly, it was.

In order to get around this problem, Lewis takes the *ancestral* of counterfactual dependence:

CT (Second Pass): C causes E if and only if there is a chain of stepwise counterfactual dependence from C to E.⁴

In *Assassination*, there is a chain of stepwise counterfactual dependence from Assassin's shooting to Victim's death, via the intermediate event of Assassin's bullet's heading towards

Victim. Assassin's bullet heading towards Victim counterfactually depends on Assassin's shooting, for, had Assassin not shot, Assassin's bullet wouldn't have headed towards Victim. In turn, given that Backup didn't shoot,⁵ Victim's death counterfactually depends on Assassin's bullet's heading towards Victim, for, had Assassin's bullet not headed towards Victim, Victim wouldn't have died. Hence, CT (Second Pass) yields the right result: Assassin's shooting caused Victim's death.

Interestingly, however, this second pass has an important drawback. As a result of analyzing causation as the ancestral of counterfactual dependence, too many things end up counting as causes, including things that, intuitively, don't make a difference to the effect. Hence the move from the first pass to the second pass is a step away from the difference-making idea, which originally served to motivate the theory. Consider, for instance, the following case:

Switch: Victim is stuck on the railroad tracks. A runaway train is hurtling down the tracks when it approaches a switch. I flip the switch, and the train turns onto a side track. However, the tracks reconverge a bit further ahead, before the place where Victim is standing. Victim dies.

Is my flipping the switch a cause of Victim's death? According to CT (Second Pass), it is, for there is a chain of stepwise counterfactual dependence from my flipping the switch to the death, via the intermediate event of the train running on the side track. This emerges as follows: The train's running on the side track counterfactually depends on my flipping the switch, for, had I not flipped the switch, the train wouldn't have run on the side track. In turn, given that the train switched tracks and thus it is no longer on the main track, Victim's death counterfactually depends on the train's running on the side track. For, if it hadn't been running on the side track, then, given that it is not running on the main track, the train would not have reached Victim and killed him.⁶ Hence, the train's running on the side track counterfactually depends on my flipping the switch, and Victim's death counterfactually depends on the train's running on the side track. So CT (Second Pass) entails that my flipping the switch caused Victim's death in *Switch*.

However, this result clashes with the difference-making idea: intuitively, my flipping the switch did not make a difference to Victim's death. To be sure, it made *some* difference, e.g., it made the death happen via the train's running on the side track. But, intuitively, this is a difference that does not matter causally.⁷

It might be objected that our causal intuitions about Switch are morally tainted: that our intuitive verdict about Switch arises from a confusion between what I am causally responsible for and what I am morally responsible for. Clearly, I am not morally responsible for the death in virtue of flipping the switch. Thus, since it is hard to keep the moral intuition apart from the causal intuition, it is natural to think that I don't cause the death by flipping the switch. But I might cause it without being morally responsible for it; after all, we cause many things that we are not morally responsible for.

However, we can address this worry by imagining a variant of the case where no moral agents are involved. Imagine, for instance, that what flipped the switch is a gust of wind, and what was lying on the track up ahead is an apple. This doesn't change our causal intuitions: it still seems that the flipping of the switch isn't a cause of the outcome – in this case, the squashing of the apple. This suggests that our intuition that my flipping the switch didn't cause the death in Switch is genuinely causal, not merely moral.⁸

We have seen that CT (in its revised formulation) does not capture the difference-making idea: it counts as causes things that, intuitively, don't make a difference to the ensuing outcomes. In particular, we have seen that, in Switch, the flip doesn't seem to make a difference to Victim's death. However, CT counts the flip as a cause of the death.

So we should look for a new way to capture the difference-making idea. I propose that we start by focusing on Switch, where CT failed. One thing that catches the eye about Switch is that, just as the *flip* doesn't make a difference to the death, the *failure to flip* wouldn't have made a difference to the death either. In other words, *whether or not* I flip the switch makes no difference to the death; it only helps to determine the route that the train takes before reaching Victim. This suggests that what

might be missing in Switch is some kind of asymmetry between my flipping the switch and my failing to flip the switch. Maybe the reason that my flipping the switch doesn't make a difference is that the contribution that it makes is not more important than the contribution that its absence would have made. Maybe, for something to be a cause, it must make a contribution that somehow *outweighs* the contribution that its absence would have made.

How could we make this thought more precise? Here is a natural suggestion. Causes are difference-makers in that the following principle, the *Causes as Difference-Makers* principle, is true:

CDM: If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

According to CDM, a cause contributes more to the effect than its absence would have contributed to it in that the absence of the cause wouldn't have been a cause itself.⁹

Consider what CDM would say about Switch. I have pointed out that, intuitively, the contributions that the flip made and that the failure to flip would have made are on a par. Hence, it is likely that, were we to count the flip as a cause, we would also have to count the failure to flip as a cause.¹⁰ But CDM doesn't allow this. So, CDM would entail that the flip *isn't* a cause. As a result, CDM would help explain our reluctance to count the flip as a cause of the death in Switch.

I will argue that CDM succeeds in capturing the difference-making idea. The rest of this paper is concerned with clarifying the content of CDM, arguing for its truth, and examining its most important consequences. First, however, let me briefly compare CDM with CT.

An important difference between CDM and CT is that, unlike CT, CDM cannot be regarded as a reductive analysis of causation, i.e., as an analysis of the concept of cause in non-causal terms. CDM is, rather, a *constraint* on theories of causation: a condition that the true analysis of causation (if there is such a thing) would have to meet. Another important difference between CDM and CT concerns the way in which

each attempts to capture the difference-making idea. We have seen that CT attempts to cash out the difference-making idea in terms of the relation of counterfactual dependence between events. According to CT, a cause makes a difference in that, if it hadn't occurred, then some event intimately related to the effect wouldn't have occurred (on the first pass, the effect itself wouldn't have occurred; on the second pass, some event in the chain of events leading to the effect wouldn't have occurred). According to CDM, a cause makes a difference by determining, not the *events* that occur in the actual and counterfactual scenarios, but the *causal relations* that obtain in the actual and counterfactual scenarios: whether a cause occurs makes a difference to whether there is a causal relation linking an event or its absence (according as the event is present or absent) to the effect.

3. EVENTS, ABSENCES, AND THE STRINGENCY OF CDM'S DEMANDS

In this section, I explain the content of CDM in more detail and I illustrate with examples. In sections 4 and 5, I lay out my argument for CDM.

I will be assuming that absences can be causes, or, as it is sometimes put, that there is *causation by omission*. This is a reasonable assumption. Intuition dictates that there is causation by omission (intuitively, the absence of rain can cause a drought, and a mother's failure to feed her child can cause the child's death), and the majority of philosophers have followed intuition on this score. This is not to say that the assumption that there is causation by omission is trouble-free; there are problems generated by letting omissions be causes, but saying that causation by omission is impossible still seems like an overreaction to such problems.¹¹ The assumption that there is causation by omission prevents CDM from being *trivially* true. If there were no causation by omission, then, clearly, it could never be the case that both an event and its absence would have caused the same effect in the scenarios where they obtain, simply because an event's absence could never cause anything.

With this assumption in place, let us examine in more detail what CDM says. On the one hand, CDM makes a claim about how the causal powers of *events* constrain the causal powers of the corresponding *absences*. Suppose that I write a letter to my mother and that makes her happy. CDM claims that, given that my writing her a letter caused her to be happy, then, had I not written a letter to my mother, my failure to write to her wouldn't have caused her to be happy.

Some words of clarification are in order. First, what is a failure? In particular, what is *my failure to write a letter to my mother*? I will adopt a common convention according to which a failure is the absence of any event of a certain type.¹² On this view, the failure to write a letter to my mother obtains just in case no event of a certain type – a writing a letter to my mother by me, at a certain time, or within a certain interval of time – occurs. More generally, if C is an event, then the absence of C obtains just in case no C-type event occurs. In a case of this sort, where C is an event, CDM claims that, if C caused E, then, had no C-type event occurred, the failure of a C-type event to occur wouldn't have caused E.

Second, how should we understand the counterfactual claims that CDM makes? In particular, how should we understand the claim: “Given that C caused E, then, *had no C-type event occurred, the failure of a C-type event to occur wouldn't have caused E*”? We can interpret it in the standard way, i.e., by appeal to possible-worlds semantics. Take the closest possible world where no C-type event occurs; that is a world where the failure of a C-type event to occur obtains. What CDM says is that the failure of a C-type event to occur doesn't cause E in that world, given that C caused E in the actual world.

Now, CDM also makes a claim about how the causal powers of *absences* constrain the causal powers of certain specific *events*. Suppose that I fail to phone Grandma on her birthday and this makes her sad. Then CDM claims that, given that my failure to phone Grandma on her birthday caused her to be sad, had I phoned her on her birthday, my phoning her wouldn't have caused her to be sad. Again, we can interpret this in terms of possible worlds. Take the closest possible world where I

phone Grandma. CDM says that, in that world, my phoning Grandma doesn't cause her to be sad. More generally, if *C* is *the failure of an event of a certain type to occur*, i.e. an absence, then CDM claims that, if *C* caused *E*, then, had an event of the relevant type been present, it wouldn't have caused *E*. This is to say, in the closest possible world, where an event of that type occurs, that event doesn't cause *E*, given that the absence of an event of that type caused *E* in the actual world.

I have explained the content of CDM in some detail. Now I will discuss its force. CDM imposes a constraint on theories of causation. How hard is it for a theory of causation to comply with CDM?

The most interesting and controversial claim that CDM makes concerns outcomes that would still have occurred in the absence of the cause, i.e., cases where the outcome doesn't counterfactually depend on the cause. In cases where the outcome does counterfactually depend on the cause, CDM is met in a trivial way. To see this, imagine that Assassin shoots and, as a result, Victim dies; however, had Assassin failed to shoot, Victim would have lived. Then it is trivially true that, whereas Assassin's shooting caused Victim's death, Assassin's failing to shoot wouldn't have caused Victim's death: it wouldn't have caused the death because the death wouldn't have occurred if Assassin had failed to shoot.

So, in order to measure the stringency of CDM's demands, we must focus on cases where the outcome doesn't counterfactually depend on the cause. The question then becomes: how hard is it to meet CDM's demands in *those* cases? I will show that it is quite hard. As a matter of fact, coming up with a theory of causation that complies with CDM is no easy task.¹³ By way of example, I will briefly review two different types of theories of causation and I will show that they both fail to meet CDM.

A first type of theory that clashes with CDM is a type of theory according to which helping to determine the causal route to an effect is sufficient for causing the effect.¹⁴ When I flip the switch in Switch, my flipping the switch makes the train run on the side track before it reaches Victim; hence, it contributes to

determining the causal route to Victim's death. So a theory of the type we are envisaging entails that my flipping the switch causes the death. However, had I *failed* to flip the switch, my failure to flip the switch would also have contributed to determining the causal route to the death, for it would have made the train run on the *main* track before it reached Victim. So a theory of this type would also entail that, had I failed to flip the switch, my failure to flip the switch would have been a cause of the death. In other words, according to a theory of this type, no matter what I did in Switch, I would have caused the death. This contradicts CDM.

A second type of theory of causation that fails to comply with CDM is a classical "regularity" view, such as Mackie's view.¹⁵ According to Mackie, C is a cause of E just in case there is a set of occurring conditions containing C that, when conjoined with some lawful regularity, entails E, and that doesn't entail E when C is removed from the set. Consider what this view would say about Assassination. In Assassination, given that Assassin shot, there is a set of occurring conditions containing the fact that Assassin shot that, when conjoined with some lawful regularity, entails the fact that Victim died, but that doesn't entail this when the fact that Assassin shot is removed from the set. This set of conditions includes, for instance, the fact that Assassin's gun was loaded, the fact that it was aimed at Victim, etc. Hence, Mackie's view would say that Assassin's shooting was a cause of Victim's death. But now imagine that Assassin hadn't shot, in which case Backup would have shot. Then there would have been a set of occurring conditions containing the fact that Assassin *didn't* shoot that, when conjoined with some lawful regularity, entails the fact that Victim died, but that doesn't entail this when the fact that Assassin didn't shoot is removed from the set.¹⁶ This set of conditions includes, for instance, the fact that Backup intended to shoot just in case Assassin didn't shoot, the fact that Backup's gun was loaded, etc. Hence, Mackie's view would entail that Assassin's shooting caused the death but, had Assassin not shot, his failure to shoot would also have caused the death. This contradicts CDM.

We have seen that CDM imposes a highly demanding constraint on theories of causation; by way of example, I have shown that two importantly different types of theories of causation fail to comply with it. In what follows, I argue that CDM succeeds in capturing the difference-making idea. Hence, if we are to respect the difference-making idea, we should reject any theory of causation that fails to comply with CDM.

4. ARGUMENT FOR CDM (PART I)

I will argue for CDM by showing that the best candidate counterexamples to CDM fail. I will look at two paradigm cases where it is most plausible to think that both an event and its absence would have caused an outcome, and I will argue that they fail. Since they fail, and since they are the best attempts at counterexamples, I will conclude that there is good reason to believe that CDM is true.

For ease of exposition, I will focus on the specific claim that CDM makes about actions and omissions of agents, but the argument is intended to have full generality. When restricted to actions and omissions of agents, CDM reads:

CDM (A/O): If an agent's acting in a certain way caused E, then, had the agent failed to act that way, the agent's failing to act that way wouldn't have caused E. Conversely, if an agent's failing to act in a certain way caused E, then, had the agent acted that way, the agent's acting that way wouldn't have caused E.

A counterexample to CDM (A/O) would have to be a case where, in the scenario where the agent acts in the relevant way, the agent's action causes an outcome E and, also, in the scenario where the agent doesn't act in the relevant way, the agent's omission causes E.

Let us single out, in particular, the following three desiderata that a counterexample to CDM (A/O) would have to meet. First, the two causes must be an action and an *omission* by an agent (as opposed to another action by the same agent). Second, the action and the omission must be *properly aligned*, that is, the omission in question must be the

failure to act in the way that caused or would have caused E. This is to say, if one of the causes is the agent's φ -ing, then the other cause must be the agent's failure to φ (as opposed to, say, the agent's failure to ψ). Third, the action and the omission must be such that, in the scenarios where they obtain, they cause the same *token* outcome, not just outcomes of the same type.¹⁷

Is it possible to find a case that meets these three desiderata? In what follows, I look at two examples. The first example I will consider is Assassination. Once again, here is the case:

Assassination: Assassin shoots Victim and, as a result, Victim dies. However, Backup is waiting in reserve. Had Assassin not shot, Backup would have, and Victim would still have died (in a very similar way, at around the same time, etc.).

One might believe that this is a counterexample to CDM (A/O) because one might reason in the following way. Assassin's shooting caused Victim's death. However, had Assassin not shot, his failure to shoot would have caused Backup to shoot, and Backup's shooting would have in turn caused Victim's death. It follows by transitivity that Assassin's failure to shoot would *also* have caused Victim's death.

Also, Assassination seems to meet the three desiderata. For, first, Backup would have acted as a result of one of Assassin's omissions. Second, the omission that Backup would have acted as a result of is precisely Assassin's failure to shoot, that is, the omission corresponding to the action that caused the death in the actual scenario. And third, Victim's death would have occurred in very much the same way if Assassin had shot or if he hadn't; hence, the death if Assassin had shot and the death if he hadn't shot are presumably the same death.

In what follows I argue that Assassination isn't a counterexample to CDM (A/O) because, while Assassin's shooting caused the death, his failing to shoot would *not* have caused the death (although it would have caused Backup to shoot, which would have caused the death). Hence, I will be arguing that we should reject the transitivity of causation at least in contexts of this type.

Let me pause here for a moment and remind you of the dialectic. This paper started out with an assumption: the assumption that the difference-making idea is worth pursuing. I said that I would be arguing that, *if* we wish to respect the difference-making idea, *then* we should endorse my view on how to cash it out. I will now put this assumption to work in the following way. I will argue that the assumption that we should respect the difference-making idea is likely to lead us to reject the transitivity of causation in some contexts and, if transitivity fails in those contexts, then it fails in Assassination. As a result, Assassination fails to be a counterexample to CDM (A/O).¹⁸

Let me start by reminding you of the following case:

Switch: Victim is stuck on the railroad tracks. A runaway train is hurtling down the tracks when it approaches a switch. I flip the switch, and the train turns onto a side track. However, the tracks reconverge a bit further ahead, before the place where Victim is standing. Victim dies.

In section 2, I pointed out that, to the extent that we wish to respect the difference-making idea, we should say that the flip isn't a cause of the death in *Switch*. For, intuitively, the flip didn't make a difference to the death. Now, if the flip isn't a cause of the death in *Switch*, then it seems that it isn't a cause in the following variant of *Switch* either:

Switch-with-Side-Track-Disconnected: Again, I am by the switch but this time I see that part of the side track is disconnected. I think that I can make the train derail by turning it onto the side track. Hence, I flip the switch and the train turns. However, Backup is waiting by the side track. When he sees that I flip the switch, he rapidly reconnects the side track. The train runs on the side track for a while, then on the main track again, and finally kills Victim.

If anything, we feel even *more reluctant* to say that my flipping the switch is a cause of Victim's death in this case, where the side track was disconnected when I flipped the switch.

Notice that my claim about *Switch-with-Side-Track-Disconnected* is a *conditional* claim. What I am suggesting is that, *if* we said that the flip isn't a cause in *Switch*, *then* we would have to say that it isn't a cause in *Switch-with-Side-Track-Disconnected*.

We might be prepared to say that the flip is a cause in Switch if, for instance, we held the view that determining the route to an event is sufficient for causing an event (which requires giving up the difference-making idea). If we held this view, then we would want to say that the flip is also a cause in Switch-with-Side-Track-Disconnected, since, in this case too, the flip determines the route to the death. My claim is only that, on the assumption that the flip isn't a cause in Switch, as the difference-making idea dictates, it is even more clearly not a cause in Switch-with-Side-Track-Disconnected.

Let me also stress that my claim about Switch-with-Side-Track-Disconnected is a purely *causal* claim and, as such, it is independent of any moral considerations. Just as we did with Switch in section 2, we can see that the intuitions about Switch-with-Side-Track-Disconnected are genuinely causal, and not merely moral, by imagining a similar scenario deprived of moral agents. Imagine, again, that what causes the switch to be flipped is a gust of wind, what reconnects the side track is a mechanism that is automatically triggered when the switch is flipped, and what is lying on the tracks, and gets squashed by the train, is an apple. Still, we feel that, if the flipping of the switch isn't a cause of the outcome in Switch, where the side track was connected all along, then it is even more clearly not a cause in Switch-with-Side-Track-Disconnected, where the side track had to be reconnected in order for the train to reach the apple.

Now, the following also seems to be true about Switch-with-Side-Track-Disconnected: my flipping the switch caused Backup to reconnect the side track, and the reconnection of the track by Backup caused, in turn, Victim's death. It is intuitively clear that my flipping the switch caused the reconnection of the track by Backup, for the flip was the event that triggered that kind of behavior in Backup: Backup was determined to reconnect the side track just in case I flipped the switch, and he acted on that decision. And it is also intuitively clear that Backup's reconnecting the side track caused Victim's death, for the death would easily have been prevented otherwise: by reconnecting the track, Backup ensured that the death happened. Hence, my flipping the switch caused Backup

to reconnect the side track, and Backup's reconnecting the side track caused Victim's death; however, on the standing assumptions about Switch and difference-making, my flipping the switch *didn't* cause Victim's death. This is to say, on the standing assumptions about Switch and difference-making, transitivity fails in Switch-with-Side-Track-Disconnected.¹⁹

I have argued that the assumption that we should respect the difference-making idea leads to the rejection of the transitivity of causation. For the difference-making idea dictates that my flipping the switch didn't cause Victim's death in Switch. Now, if my flipping the switch didn't cause Victim's death in Switch, then it probably didn't cause it in Switch-with-Side-Track-Disconnected either. But then it seems that we should say that, in Switch-with-Side-Track-Disconnected, my flipping the switch caused Backup to reconnect the side track, which caused the death, but my flipping the switch didn't cause the death. In what follows, I argue that the scenario where Assassin fails to shoot in Assassination is on a par with the scenario where I flip the switch in Switch-with-Side-Track-Disconnected. Hence, if transitivity fails in Switch-with-Side-Track-Disconnected, when I flip the switch, it also fails in Assassination when Assassin fails to shoot.

I have pointed out that my flipping the switch is even more clearly *not* a cause of Victim's death in Switch-with-Side-Track-Disconnected than in the original case, Switch. Why is this? Intuitively, this is because my flipping the switch only made it *more difficult* for the death to happen, by calling for Backup's intervention. Given that I flipped the switch, Backup had to intervene or else the death wouldn't have happened, while, had I not flipped the switch, the death would have occurred much more easily, without the need for Backup's intervention.²⁰ Now, I will argue that the relation between Assassin's failure to shoot and Victim's death in Assassination is significantly similar to the relation between my flipping the switch and Victim's death in Switch-with-Side-Track-Disconnected. Hence, if my flipping the switch does not cause the death in Switch-with-Side-Track-Disconnected, Assassin's failure to shoot does not cause the death in Assassination either.

To see this, imagine that Assassin failed to shoot in Assassination. So Backup shot, and Victim died. Then, just as in Switch-with-Side-Track-Disconnected, Assassin's failure to shoot only made it more difficult for Victim's death to happen, by calling for Backup's intervention. Given that Assassin didn't shoot, Backup had to intervene or else the death wouldn't have happened, while, had Assassin shot, the death would have occurred much more easily, without the need for Backup's intervention. This suggests that the same reasons that should lead us to reject transitivity in Switch-with-Side-Track-Disconnected (in the scenario where I flip the switch) should also lead us to reject transitivity in Assassination (in the scenario where Assassin fails to shoot). They should lead us to say that, while Assassin's failure to shoot would have caused Backup to shoot, and while Backup's shooting would have caused Victim's death, Assassin's failure to shoot would *not* have caused Victim's death. If so, Assassination fails to be a counterexample to CDM (A/O) because it is not true that *both* Assassin's shooting *and* Assassin's failing to shoot would have caused Victim's death.

My diagnosis of Assassination can be generalized to cases of the following sort. An agent's action and the corresponding omission would both have been followed by a certain outcome. The agent's action is the sort of action that can issue in the outcome without the aid of any backup mechanism. The agent's omission, by contrast, is the sort of omission that can only issue in the outcome if a backup mechanism intervenes. Given that the omission calls for a backup mechanism that the action doesn't call for, it seems wrong to count the omission as a cause of the outcome. Hence, it is not the case that both the action and the omission would have caused the same outcome. Hence, cases of this type aren't counterexamples to CDM (A/O).

In cases of the type that we have just examined, one of the candidates for being a cause, the agent's action, has an initial causal advantage over the other candidate, the agent's omission, and thus, it is a better *prima facie* candidate for being a cause. But, what about cases where the two candidates are intuitively *on a par*? That is, what about cases where neither

candidate is a better *prima facie* candidate for being a cause? Couldn't cases of this type be counterexamples to CDM (A/O)? I turn to a case of this type in section 5.

5. ARGUMENT FOR CDM (PART II)

Here is such a case:

Two-Assassins: I hired two assassins and I gave them the following instructions. Assassin 1 is to shoot Victim just in case I nod at t . Assassin 2 is to shoot Victim just in case I fail to nod at t . As a matter of fact, I nod at t , Assassin 1 shoots and Victim dies.

Someone might want to say that this is a counterexample to CDM (A/O) for reasons parallel to those mentioned in our discussion of Assassination. Namely, my nodding caused Assassin 1 to shoot, which caused the death; hence, it is tempting to say that my nodding caused the death. Similarly, my failure to nod would have caused Assassin 2 to shoot, which would have caused the death; hence, it is tempting to say that my failure to nod would also have caused the death.

Also, *Two-Assassins* seems to meet the desiderata from section 4. First, the two candidate causes are an agent's action and an agent's omission. Second, the action and the omission are properly aligned: Assassin 2 would have shot just in case I failed to nod, where my nodding is precisely that which made Assassin 1 shoot in the actual scenario. Third, we can imagine that Victim would have died in very much the same way, and around the same time, if Assassin 2 had shot instead of Assassin 1, so that the death that Victim would have undergone if Assassin 2 had shot would have been the same death as the one that he underwent given that Assassin 1 shot.

Finally, *Two-Assassins* is a case where the agent's action and the agent's omission are intuitively on a par with respect to their causal powers: it seems that one of them is a cause just in case the other is a cause. Hence, my argument against Assassination from the last section doesn't apply to *Two-Assassins*.

Still, I will argue that *Two-Assassins* fails to be a counterexample to CDM (A/O) because (on the standing assumptions

about Switch and difference-making) transitivity fails in this case as well. However, my diagnosis of Two-Assassins will differ from that of Assassination, in the following way. I have claimed that, in Assassination, while Assassin's shooting caused the death, his failure to shoot wouldn't have caused it. By contrast, I will claim that, in Two-Assassins, *neither* my nodding *nor* my failure to nod would have caused the death. This is to say, I will argue that, in a case where the agent's action and the omission are intuitively on a par, neither is a cause of the outcome.

Again, my argument will be based on an analogy with a variant of Switch. The variant that we need now is one where, not just one, but *the two* tracks are initially disconnected:

Switch-with-Both-Tracks-Disconnected: This time, both of the tracks are disconnected after the switch. However, there is one assassin next to each track. If I don't flip the switch, Assassin 1 will reconnect the main track and Victim will die. If I flip the switch, Assassin 2 will reconnect the side track and Victim will die.

Suppose that I flip the switch. Consequently, the train turns onto the side track, which Assassin 2 rapidly reconnects, then the tracks reconverge, and Victim dies. Did my flip cause the death? If it didn't cause it in the original case, Switch, it seems that it didn't cause it in this case either. Intuitively, in neither case did the flip make a difference to the death, for the actual scenario and the scenario where I don't flip the switch are relevantly parallel: whereas Assassin 2 reconnects the track in the actual case, Assassin 1 reconnects the track in the case where I don't flip the switch. Hence, it seems that, if the flip didn't cause the death in Switch, then it didn't cause the death in Switch-with-Both-Tracks-Disconnected either. But it is clear that, in Switch-with-Both-Tracks-Disconnected, my flipping the switch caused Assassin 2 to reconnect the side track, which in turn caused the death. Hence, on the standing assumptions about Switch and difference-making, transitivity fails in Switch-with-Both-Tracks-Disconnected if I flip the switch.

Alternatively, suppose that I *don't* flip the switch. The train then continues along the main track, which Assassin 1 rapidly

reconnects, and Victim dies. Again, it seems that, if my failure to flip the switch didn't cause the death in Switch, then it didn't cause it here either, even though it caused Assassin 1 to reconnect the track, which in turn caused the death. Thus, on the standing assumptions about Switch and difference-making, transitivity fails *both* if I flip the switch *and* if I don't flip the switch.

Now, Two-Assassins strikes me as on a par with Switch-with-Both-Tracks-Disconnected. Just as neither my flipping the switch nor my failing to flip the switch would have made a difference to Victim's death in Switch-with-Both-Tracks-Disconnected, it seems that neither my nodding nor my failing to nod would have made a difference to Victim's death in Two-Assassins. For, in Two-Assassins too, the scenario where I nod and the scenario where I fail to nod are relevantly parallel: whereas Assassin 1 shoots in the case where I nod, Assassin 2 shoots in the case where I don't nod. Thus, if neither the flip nor the failure to flip would have caused the death in Switch-with-Both-Tracks-Disconnected, then, similarly, neither my nodding nor my failure to nod would have caused the death in Two-Assassins. My nodding wouldn't have caused the death, even though it would have caused Assassin 1 to shoot, which would have caused the death. And my failure to nod wouldn't have caused the death, even though it would have caused Assassin 2 to shoot, which would have caused the death.

Now, one might find this puzzling. I hired Assassin 1 and gave him the instruction to shoot just in case I nodded. How can I say, then, that my nodding wouldn't have caused the death? Similarly, I hired Assassin 2 and gave him the instruction to shoot just in case I didn't nod. How can I say, then, that my failure to nod wouldn't have caused the death?

To see that this isn't a problem, imagine that I also hired the two assassins standing by the tracks in Switch-with-Both-Tracks-Disconnected. Thus, imagine that I gave Assassin 1 the instruction to reconnect the main track in case I didn't flip the switch, and Assassin 2 the instruction to reconnect the side track in case I flipped the switch. This doesn't change the verdict about the causal powers of my flipping the switch, or of my failure to flip the switch. To be sure, if I hired the two assassins,

I caused the death. But I caused it by *hiring the two assassins*, not by flipping the switch or failing to flip it.²¹ By hiring the two assassins and giving them the instructions that I gave them, I made sure that the death happened. But I also made sure that, at the time in which I had to decide whether to flip the switch or not, what I decided to do then couldn't make a difference. This is to say, I made sure that nothing I did or failed to do at that moment could count as a cause of the death.²²

Similarly, my claim is that, in Two-Assassins, I caused Victim's death but not in virtue of nodding, or failing to nod. I caused the death by hiring the two assassins and giving them the specific instructions that I gave them. By hiring the assassins and giving them those instructions, I made sure that the death happened, but I also made sure that, at the time in which I had to decide whether to nod or not, what I did then couldn't make a difference. This is to say, I made sure that nothing I did or failed to do at that moment could count as a cause of the death.

Let me sum up the results of this section and the preceding section. I have argued that, on the assumption that we should respect the difference-making idea, it follows that two main attempts at counterexamples to CDM (A/O), Assassination and Two-Assassins, fail. I first argued that Assassination isn't a counterexample to CDM (A/O), for, while the agent's action is a cause of the outcome, the agent's omission wouldn't have been a cause of the outcome. Then I pointed out that Assassination is a case where, intuitively, one of the candidate causes, the agent's action, has an initial advantage over the other, the agent's omission. So the natural reaction was to look instead for a case where the action and the omission are intuitively on a par with respect to their causal powers and to see whether a case of that sort has a better chance of being a counterexample to CDM (A/O). This is how we arrived at Two-Assassins. I argued, however, that, as a result of making the action and the omission equally good candidate causes, as in Two-Assassins, it turns out that *neither* is a cause, not that both are. This is to say, by depriving the candidate causes of any initial advantage over each other, we deprive them of causal power altogether. Hence, Two-Assassins also fails to be a counterexample to CDM (A/O).

My argument for CDM (A/O) takes, then, the following form. Possible counterexamples to CDM (A/O) can be grouped into two main classes: the class of cases where the action and the omission aren't intuitively on a par and the class of cases where the action and the omission are intuitively on a par. My discussion of Assassination suggests that the cases in the *first* class fail because *only one* of the candidate causes is a genuine cause. In turn, my discussion of Two-Assassins suggests that the cases in the *second* class fail because *neither* of the candidate causes is a genuine cause. Since any alleged counterexample will fall into one of the two classes, and since the cases I have discussed seem to be representative of their class, I conclude that there is good reason to believe that CDM (A/O) is true. More precisely, I conclude that there is good reason to believe that CDM (A/O) succeeds in capturing the difference-making idea in the case of actions and omissions of agents.

Finally, my focus on actions and omissions of agents was only for simplicity. In principle, it should be possible to use the same style of argument to show that the general claim, CDM, is true. I conclude that CDM succeeds in cashing out the difference-making idea: if causes are difference-makers, it is in virtue of the fact that events and their absences would not have caused the same effects.

6. IMPLICATIONS FOR MORAL RESPONSIBILITY

In this section, I draw attention to some important results that CDM has for issues in moral responsibility. I argue, in particular, that CDM achieves a nice fit between the concepts of causation and moral responsibility.

One way in which this emerges is as follows. Ordinarily, we regard ourselves as morally responsible for the (foreseeable) consequences of what we do or fail to do. Intuitively, this seems to be because, ordinarily, we regard ourselves as having a *choice* whether to cause those consequences. I say "ordinarily," because there are some *extraordinary* circumstances where this is not the case. Notably, if we are coerced to

behave in certain ways, or if we act under the influence of some powerful drug, then we might not have a choice whether to cause the ensuing consequences and thus we might not be responsible for them. But these cases are extraordinary in that they are cases where we lack control of the *actions and omissions* that issue in those consequences and, correspondingly, they are cases where we do not have a choice whether to cause those consequences.

Now, according to CDM, and, in particular, according to CDM (A/O), *whenever* we have a choice whether to act or fail to act in certain ways, we *thereby* have a choice whether to cause the ensuing consequences. According to CDM (A/O), it simply couldn't be that, both by acting and by failing to act, I would be causing the same consequences. Hence, CDM (A/O) suitably fits the way in which we ordinarily think of ourselves as responsible for the consequences of our actions and omissions.

By contrast, imagine what we would have to say if we rejected CDM (A/O). If we rejected CDM (A/O), then we would have to say that, on some occasions, both acting in certain ways and failing to act in those ways would cause the same outcomes. Then, on those occasions, we wouldn't have a choice whether to cause those outcomes. For, in those cases, regardless of what we did (were we to act in certain ways or were we to fail to act in those ways), we would be causing those outcomes. Moreover, in those cases, we wouldn't have a choice whether to cause certain outcomes *even if we happened to be in complete control of the actions and omissions that caused those outcomes*. But, as I have pointed out, barring exceptions in which we lack control of our own actions and omissions, we tend to regard ourselves as responsible for the consequences of our actions and omissions because we tend to regard ourselves as having a choice whether to cause those consequences. Hence, rejecting CDM (A/O) would clash with the ordinary way in which we think of ourselves as responsible for the consequences of our actions and omissions.

Let me illustrate this point with an example, before moving on to my last remark. Take Assassination. Suppose that I have the choice between shooting and failing to shoot. Intuitively, I

then have a choice whether to cause Victim's death. Regardless of whether Backup shoots, and regardless of whether Backup shoots as a result of my failing to shoot or because he was going to shoot anyway, the intuitive thought is that, if I don't shoot, I don't cause the death, and therefore I am not responsible for the death. Now, someone might be prepared to give up this thought, upon realizing that it requires rejecting the transitivity of causation. What I am suggesting is that this would come at a cost: the cost of giving up the ordinary way in which we regard ourselves as responsible for the consequences of our actions and omissions. CDM takes the opposite route: it allows for the intransitivity of causation, but it accommodates the ordinary way in which we regard ourselves as responsible for the consequences of our actions and omissions.

Finally, I will suggest that CDM is particularly helpful in accounting for the lack of moral responsibility of agents in some cases of moral luck. Briefly, a case of moral (good) luck is a case where an agent that behaves in a morally wrong way doesn't come out responsible for a harm thanks to the obtaining of some circumstances that are outside of the agent's control.²³ Here is a case of moral luck with respect to which CDM can prove particularly useful:

Switch-with-Main-Track-Unexpectedly-Connected: Again, Victim is trapped on the tracks. I want Victim to die, and I have reason to believe that the main track is disconnected. So, thinking that the train will derail if it continues on the main track, I flip the switch. As it turns out, however, the main track has never been disconnected. As a result of my flipping the switch, the train turns onto the side track, but then the tracks reconverge and the train hits Victim.

Intuitively, this is a case of moral luck because, even if I acted wrongly in flipping the switch, I am not responsible for Victim's death (I might be responsible for intending to cause his death, for trying to cause his death, etc., but not for the death itself). I *thought* that I would cause Victim's death by flipping the switch, and I *intended* to cause the death by flipping the switch. However, even if the death did occur, it seems that, given that the main track was connected all along, my flipping the switch

did not cause the death. Switch-with-Main-Track-Unexpectedly-Connected only differs from the original case, Switch, in what I *thought* was the case, not in what was actually the case. Hence, if my flipping the switch does not cause the death in Switch, it does not cause it in Switch-with-Main-Track-Unexpectedly-Connected either.

Now, as I have suggested, CDM (together with the observation that the contribution that the flip made to the death is on a par with the contribution that the failure to flip would have made to the death) entails that the flip did not cause the death in Switch. For the same reason, CDM entails that the flip did not cause the death in Switch-with-Main-Track-Unexpectedly-Connected. As a result, CDM helps to explain my moral luck in this case.

By contrast, many theories of causation entail that my flipping the switch *did* cause the death in Switch-with-Main-Track-Unexpectedly-Reconnected. By way of example, note that my flipping the switch helped to determine the causal route to the death, by determining the path that the train takes before reaching the person. Hence, in particular, theories of causation according to which helping to determine the causal route to an outcome is sufficient for causing the outcome entail that my flipping the switch *is* a cause of the death. As a result, these theories fail to account for my moral luck in cases of this type.²⁴

I conclude that, not only does CDM succeed in capturing the difference-making idea, but it also yields a concept of cause that has the seemingly right kinds of connections to moral concepts, such as the concept of moral responsibility. This is an attractive feature of CDM.

NOTES

¹ Lewis (1986a), pp. 160–161.

² In recent years, Lewis's theory gave rise to an array of revisions and adjustments, all of which attempt to analyze the concept of causation, ultimately, in terms of counterfactual dependence between events. Two examples are McDermott (1995) and Lewis himself in his later work, Lewis (2000). Henceforth, by "Lewis's counterfactual theory" I mean the theory developed in his (1986a).

³ This theory of causation assumes determinism. I will not be concerned with the possibility of indeterministic causation in this paper.

⁴ Lewis (1986a), p. 167.

⁵ According to Lewis, the standard contexts of evaluation of counterfactuals are not “backtracking.” Thus, in considering a counterfactual of the form “If C hadn’t occurred, E wouldn’t have occurred,” we must hold fixed as much of what happened before C as possible. see Lewis (1986b).

⁶ Where would have the train gone, then? This depends on the details of our theory of counterfactuals. Maybe it would have derailed, or it would have miraculously vanished. Either way, it wouldn’t have reached Victim, given that Victim could only be reached via one of the tracks.

⁷ Lewis’s most recent attempt, the “causation as influence” view (in Lewis (2000)), has the same kind of problem. For there Lewis analyzes causation as the ancestral of the influence relation. As a result, he counts as causes things that, intuitively, make no difference to the effects.

⁸ For attempts to rescue the intuition that my flipping the switch isn’t a cause, see Rowe (1989); Yablo (2002) and (2004), I do not wish to suggest, however, that these authors would agree with the proposal I will offer shortly.

⁹ I intend this to apply to both “positive” and “negative” causes. For instance, C could be an omission, in which case the absence of C would be an action (more on this below).

¹⁰ On the assumption that there is causation by omission. I discuss this assumption shortly.

¹¹ Causation by omission would be a problem if, for instance, one believed that the causal relata are events. For, on many views, omissions aren’t events. On the other hand, allowing for the possibility of causation by omission helps to preserve the important connection that seems to exist between causation and moral responsibility. If a mother doesn’t feed her baby, it seems that she is morally responsible for the baby’s death in virtue of having caused his death by not feeding him. But this is causation by omission.

¹² See, e.g., Lewis (1986a), p. 189.

¹³ A theory that clearly meets CDM is CT (First Pass). On this view, if C is a cause of E, then E wouldn’t have occurred in C’s absence; hence, it is clear that the absence of C wouldn’t have caused E. However, as we have seen, there are obvious counterexamples to this view. As a result, no one seems to hold it.

¹⁴ An example of a theory of this type is CT (Second Pass).

¹⁵ Mackie (1993). See, in particular, the example on p. 43, where Mackie seems to suggest that his view has the consequence that it is possible for the fact that an event occurs to cause an outcome when the fact that the event doesn’t occur would also have caused it. (Notice that, for Mackie, the causal relata are facts, not events. But nothing essential hangs on this.)

¹⁶ This last part is true because Backup wouldn’t have shot *unless* he saw that Assassin didn’t shoot.

¹⁷ These desiderata help to distinguish CDM from clearly false theses in the vicinity. For instance, the claim that there cannot be more than one way to cause an outcome, and the claim that it is impossible for an action and the corresponding omission to cause outcomes of the same type.

¹⁸ The assumption that we should respect the difference-making idea plays an important role in my argument because it is not easy to argue against the transitivity of causation. Merely pointing to seeming counterexamples to transitivity does not seem to be enough, for the view that causation is transitive seems to be deeply entrenched in our way of thinking. When we look for the causes of a given event, we often proceed by tracing a causal chain back to earlier events and concluding that those earlier events are causes of the later event. This method assumes that causation is transitive. For discussion of transitivity, see Hall (2000); Hitchcock (2001); Paul (2000); Yablo (2002) and (2004).

¹⁹ For some people, the failure of transitivity would arise earlier in my argument, in Switch itself. This would be so if we believed that, while my flipping the switch didn't cause the death, it caused the train to run on the side track, which caused the death. However, we needn't say this about Switch. Maybe what my flipping the switch caused wasn't what caused the death. Maybe what caused the death was the event of the train's running towards Victim, and what my flipping the switch caused was the fact that such event had a certain feature, i.e., the fact that it happened on the *side* track. L. Paul rebuts some alleged counterexamples to transitivity in this way in Paul (2000). Notice, however, that Switch-with-Side-Track-Disconnected doesn't seem to be open to the same treatment: it seems clear that my flipping the switch caused Backup to reconnect the track, and that this caused the death.

²⁰ Here is another example in the same vein. While swimming in the sea, a child is attacked by a shark. The child is then rushed to a hospital, where he is treated for a few days, until his wounds heal. Intuitively, the shark attack caused the medical treatment, the medical treatment caused the child's good health, but the shark attack did *not* cause the good health. This is so because, intuitively, the shark attack only made it more difficult for the child's good health to ensue, given that it introduced the need for the medical treatment.

²¹ How did my hiring the two assassins cause the death? By starting a causal route to the death, which included the reconnection of the side track by the assassin on that track. Rejecting transitivity is consistent with saying this. More generally, rejecting transitivity is consistent with saying that, in order for C to be a non-immediate cause of E, there must be an intermediary, D, that is caused by C and that causes E.

²² What if I could call the whole thing off by, say, waving my hand in a particular way? Then my failure to wave my hand in that way would have been a cause of Victim's death. Still, my flipping the switch or my failing to flip the switch wouldn't have been a cause of the death.

²³ See Nagel (1979), chap. 3, and Williams (1981), chap. 2. All of Nagel's and Williams' examples are cases where the agent isn't responsible for a harm that doesn't occur but could easily have occurred. By contrast, I will focus on cases where the harm does occur but the agent doesn't cause it.

²⁴ Typically, intentionally causing a harm (in a "non-deviant" way) is taken to be sufficient for being morally responsible for the harm. (See, e.g., Feinberg (1970).) If this is so, any theory of causation that entails that I caused the death will have serious trouble explaining my moral luck. For, if I caused the harm by flipping the switch, I did it intentionally (and in a "non-deviant" way, i.e., by making the train turn onto the side track, as I intended). As a result, if we said that my flipping the switch caused the death, it is likely that we would have to revise our views on moral responsibility.

REFERENCES

- Feinberg, J. (1970): 'Sua Culpa', in *Doing and Deserving: Essays in the Theory of Responsibility*, Princeton: Princeton U.P., pp. 187–221.
- Hall, N. (2000): 'Causation and the Price of Transitivity', *Journal of Philosophy* 97(4), 198–222.
- Hitchcock, C. (2001): 'The Intransitivity of Causation Revealed in Equations and Graphs', *Journal of Philosophy* 98(6), 273–299.
- Lewis, D. (1986a): 'Causation', reprinted in *Philosophical Papers, Vol. II*, New York: Oxford U.P., pp. 159–213.
- Lewis, D. (1986b): 'Counterfactual Dependence and Time's Arrow', reprinted in *Philosophical Papers, Vol. II*, New York: Oxford U.P., pp. 32–66.
- Lewis, D. (2000): 'Causation as Influence', *Journal of Philosophy* 97(4), 182–197.
- Mackie, J. (1993): 'Causes and Conditions', reprinted in E. Sosa and M. Tooley (eds.), *Causation* (pp. 33–55), New York: Oxford U.P.
- McDermott, M. (1995): 'Redundant Causation', *British Journal for the Philosophy of Science* 46, 523–544.
- Nagel, T. (1979): *Mortal Questions*, New York: Cambridge U.P.
- Paul, L. (2000): 'Aspect Causation', *Journal of Philosophy* 97(4), 235–256.
- Rowe, W. (1989): 'Causing and Being Responsible for What is Inevitable', *American Philosophical Quarterly* 26(2), 153–159.
- Williams, B. (1981): *Moral Luck*, Cambridge: Cambridge U.P.
- Yablo, S. (2002): 'De Facto Dependence', *Journal of Philosophy* 99, 130–148.
- Yablo, S. (2004): 'Advertisement for a Sketch of an Outline of a Proto-Theory of Causation', in J. Collins, N. Hall and L. Paul (eds.), *Causation and Conditionals*, Cambridge: M.I.T. Press.

Department of Philosophy
University of Wisconsin-Madison
5185 Helen C. White Hall
600 North Park St.
Madison, WI 53706
E-mail: acsartorio@wisc.edu