

The Prince of Wales Problem for Counterfactual Theories of Causation *

1. Introduction

In 1992, as part of a larger charitable campaign, the Prince of Wales (Prince Charles, Queen Elizabeth's older son and heir) launched a line of organic food products called "Prince's Duchy Originals" (www.duchyoriginals.com). The first product that went on sale was an oat cookie: "the oaten biscuit." Since then the oaten biscuit has been joined by hundreds of other products and Duchy Originals has become one of the leading organic food brands in the UK. Presumably, the Prince of Wales is very proud of his Duchy Originals products, and of the oaten biscuits in particular. Let's imagine that he is so proud of the biscuits that he eats them regularly. Also, let's imagine that one day Queen Elizabeth asks the Prince to water her plant. As she explains to him, she'll be gone for the day and the plant needs to be watered every afternoon. But the Prince decides not to water the plant. Instead of watering it, he spends his afternoon savoring some oaten biscuits, and the plant dies.

What caused the plant's death? If you were to ask the Queen, she would presumably say: the Prince, plus some "natural causes" (including the fact that the plant was particularly delicate and needed intensive watering). Now, in virtue of what could the Prince be a cause of the plant's death? When we say that an agent caused some event in the world, we typically mean to say that there is something that the agent did, or something that the agent failed to do, which caused the outcome. There are several things that the Prince did and failed to do that afternoon: he ate some oaten biscuits, he read the newspaper, he scratched his nose, he didn't phone a friend, he didn't watch TV, he didn't water the Queen's plant, etc. Among these, we

* For helpful comments, thanks to Juan Comesaña, Dan Hausman, Allan Hazlett, Patrick Monaghan, Elliott Sober, and audiences at the University of Iowa and the 2009 Pacific APA.

clearly want to say that his not watering the plant is relevant to the plant's death: the plant died because he didn't water it. Under slightly different circumstances, some of the things he *did* would also be relevant. Imagine, for instance, that the oaten biscuits are so amazingly good that they induce some kind of psychological trance that makes you forget any obligations that you might have. So maybe the Prince was determined to water the plant until he ate the biscuits, at which time he forgot all about it. In that scenario his eating the biscuits would also be a cause of the plant's death. But note that, even in that case, the Prince's contribution to the plant's death is ultimately "negative" in nature. For his eating the biscuits causes the plant's death *by means of* causing the Prince's subsequent failure to water the plant. At the end of the day, the plant still dies because of something that the Prince *doesn't* do: it dies because he doesn't water it.

Scenarios of this kind suggest that omissions, and absences in general, can be causes, and that our reconstructions of the causal histories of the outcomes are somehow flawed if they don't include the omissions of agents or the absences of certain events but instead include only "positive" causes. For, again, in these cases, the outcomes seem to happen, at least partly, because of something that someone doesn't do, or because of something that doesn't happen, not (just) because of something that someone does, or because of something that actually happens. I will call the apparent failure of positive causes to adequately account for the outcome's occurrence in these cases "the inadequacy fact about positive causes." The inadequacy fact about positive causes is an important motivation for accepting negative causes.

Now, assuming that we want to make room for negative causes, how could we make sense of omissions and absences being causes? A natural thought is to appeal to the notion of

counterfactual dependence. We can say that the Prince's not watering the plant is a cause of the plant's death because the plant's death counterfactually depends on the Prince's failure to water it: had his failure to water the plant not occurred (i.e. had he watered the plant), the plant wouldn't have died. In other words: in the closest possible world(s) where the Prince waters the plant, the plant doesn't die. Counterfactual theories of causation claim that the causal facts are grounded in facts about counterfactual dependence. On these views, causes are "difference-makers" with respect to their effects in that effects (at least typically) counterfactually depend on their causes (Lewis 1986a).

Now, this idea has to be refined in two kinds of ways.¹ First, as cases of "preemption" suggest, sometimes effects don't counterfactually depend on their causes. For example, an assassin can cause his victim's death even if the death would still have happened if he hadn't shot him, given that a backup assassin would then have shot the victim himself. This suggests that counterfactual dependence is not necessary for causation. At least originally, Lewis thought that we can sidestep this problem by taking causation to be, not simple counterfactual dependence, but the ancestral of counterfactual dependence. Second, counterfactual dependence also doesn't seem to be sufficient for causation: some counterfactual dependencies track "tighter," non-causal connections, such as logical and mereological relations. For example, my writing the word 'cat' counterfactually depends on my writing the letter 'c' but my writing 'c' isn't a cause of my writing 'cat' (Kim 1973). So the relevant concept of counterfactual dependence would have to be circumscribed accordingly. Lewis does this by setting constraints on potential causes and effects. On Lewis' view, a necessary condition for C to cause E is that C and E be fully "distinct," where C and E are not fully

¹ For a more extensive discussion of these problems, see Lewis 1986a, 1986b, and 1986c.

distinct if, for example, one is part of the other. Also, some counterfactual conditionals express counterfactual relations that aren't causal because they are *backtracking*—as when I say “If my friend had invited me to his birthday party today, then we wouldn't have had a fight yesterday.” Lewis' suggestion is that we should restrict our focus to ordinary or standard contexts, in which backtracking counterfactuals aren't true.

In spite of these problems, the claim that counterfactual views have at least identified a *sufficient* condition for causation once counterfactual dependence has been restricted in these ways has seemed quite plausible to people. In particular, at least in recent times, it has seemed much more plausible than the converse claim that counterfactual dependence, or something close to it, is *necessary* for causation (the consensus seems to be that counterfactual theories have a really hard time addressing the preemption problem). From now on I will focus on the sufficiency claim only. I'll call it “the counterfactual criterion”:

(CC) If there is counterfactual dependence of the ordinary (non-backtracking) kind between C and E, and if C and E are fully “distinct” (e.g., they are not logically or mereologically related), then C is a cause of E.

CC seems to be initially plausible: if E counterfactually depends on C, then C is a difference-maker with respect to E—it makes the difference between E's occurring and E's not occurring—and so (if the counterfactual dependence is of the ordinary kind and if C and E are

fully distinct) it is plausible to think that C is one of the things (among potentially multiple things) that causally contributed to E's occurrence.²

Given CC's initial plausibility, an advantage of counterfactual theories seems to be that they have the basic resources to accommodate causation by omission, which many other theories lack.³ For example, theories according to which a causal relation requires the transfer of some physical quantity, like energy or momentum (Salmon 1994), or any other kind of physical interaction, don't have the resources to do this. For there is no physical interaction between the Prince and the plant in virtue of which he caused the plant's death. Counterfactual dependence, by contrast, doesn't require the existence of physical interaction: on the basis of CC, we can say that the Prince caused the plant's death even if he never physically interacted with it. So the ability to accommodate causation by omission appears to be at least a *prima facie* advantage of counterfactual theories over theories that don't allow for this type of causation.⁴

I will argue that this is a misconception. I will argue that, despite appearances to the contrary, the ability to accommodate causation by omission is not a *prima facie* advantage of counterfactual views, at least to the extent that we take the main motivation for believing in causation by omission to be the inadequacy fact about positive causes (as I am assuming we

² A few people would reject CC. Thomson would reject it on the grounds that it entails, e.g., that a person's birth always causes that person's death (Thomson 2003, section 11). Lewis responds to an objection of this kind in his 2002, p. 101. Hall (2004) also rejects CC, but only for one concept of causation (he thinks there are two concepts of causation: the "productive" concept and the "dependence" concept). Yablo (1992) also rejects CC. Yablo's view is particularly relevant for our purposes; I discuss it in section 4 below.

³ To say that counterfactual theories have the basic resources to account for causation by omission is not to say that there aren't problems arising from the intangibility and elusive nature of omissions. Lewis, in particular, had trouble fitting causation by omission into his general picture of causation and events (see Lewis 1986a, Postscript D). He takes on the topic of causation by omission again in Lewis 2004.

⁴ Of course, this is consistent with there being good overall reasons for rejecting counterfactual theories (as well as for rejecting the possibility of causation by omission). For a recent debate about the legitimacy of causation by omission, see Dowe 2004 and Schaffer 2004 (Schaffer is in favor, and Dowe is opposed).

do). For I will argue that, even if omissions are causes, and even if counterfactual views can accommodate causation by omission, those views still fail to respect the inadequacy fact about positive causes. Although my main focus will be counterfactual theories of *causation*, in the final section I will suggest that the arguments of this paper apply, more generally, to theories that attempt to account for the contribution of agents' omissions in counterfactual terms, regardless of whether this is a causal contribution or not.

2. The problem of unwanted positive causes

Let's start by drawing attention to a familiar problem that arises for counterfactual views of causation by omission. As several people have pointed out, the ability that counterfactual theories have to account for causation by omission seems to backfire (Schaffer 2000, Thomson 2003, Beebe 2004, McGrath 2005). Consider the plant in my backyard, which also just died from lack of water. Had I watered it, it would have survived; so, given that I didn't, CC entails that I caused it to die by failing to water it (note that the counterfactual dependence that exists between my not watering the plant and the plant's death is not backtracking, logical, mereological, etc.). But if, say, the *Queen of England* had watered it, it *also* would have survived (and, again, the counterfactual dependence between the Queen's not watering my plant and the plant's death is of the right kind). So it follows from CC that the Queen also caused it to die, by failing to water it! But it seems wrong to say that the Queen caused my plant to die. In other words, the problem is that, in accounting for causation by omission in counterfactual terms, we go from *too few* causes to *too many* causes. Let's call this apparent problem for counterfactual theories of causation "the Queen of England problem."⁵

⁵ I gave the problem this label in Sartorio 2004.

Now, although this is a surprising result, I don't consider the Queen of England problem to be a genuine problem. For, on reflection, there seems to be no metaphysically relevant difference between the Queen and me in virtue of which we can say that I'm a cause of my plant's death but the Queen isn't. Even if the plant is mine and not the Queen's, even if I regularly water it and the Queen doesn't, etc., these don't seem to be grounds for a genuine *causal* difference, but only for a difference in what it is reasonable to expect each one of us will do. So perhaps CC is right and I am a cause of the plant's death and so is the Queen. Still, I am a much more *salient* cause than the Queen, given that I was expected to water it. This might be sufficient to explain our initial reluctance to regard the Queen as a cause.

At any rate, I will assume that this is a satisfactory answer to the Queen of England problem. But I will argue that there is another problem that arises for CC concerning omissions and that is, unlike the Queen of England problem, a genuine and serious problem. I will also give it a royal name, *the Prince of Wales problem*, for it can be seen as a close relative of the Queen of England's problem, in fact, as its "successor." The problem is, again, that by embracing CC we go from too few causes to too many causes. But, in this case, by contrast with the Queen of England problem, there is no good reason to think that those things that CC entails are causes are really causes. Moreover, in this case the unwanted causes that CC lets in are of both kinds: negative and positive. So the counterfactual strategy *truly* backfires, and it does so *doubly*. In what follows, I discuss the problem of unwanted positive causes (I discuss the opposite problem—the problem of unwanted negative causes—in the next section).

Let's draw a basic but helpful distinction between different ways in which agents can causally contribute to outcomes in the world. As we have mentioned, on the face of it agents

can causally contribute to outcomes either by acting in certain ways or by failing to act in certain ways. Let's say that an agent's contribution is *positive* when the agent causes the outcome in virtue of having acted in a certain way. For example, an agent's contribution is positive when the agent does something that is part of a physical process leading to the outcome (as when someone poisons a plant and the plant dies as a result). On the other hand, let's say that an agent's contribution is *negative* when the agent causes the outcome in virtue of having failed to act in a certain way. For example, an agent's contribution is negative when the agent fails to interfere with an existing physical process that eventually leads to the outcome (as when someone fails to water a plant and the plant dies). Finally, let's say that an agent's contribution is *mixed* when it's both positive and negative. Some contributions are mixed in that they contain successive positive and negative components (as in the amnesia-inducing-biscuits scenario imagined in the previous section: the Prince's eating the biscuits results in the plant's death by means of resulting in the Prince's failure to water the plant). Other causal contributions are mixed in that they contain simultaneous or overlapping positive and negative components (if I push a boulder over your head and I fail to warn you about it, I am responsible for the fact that the boulder is falling over your head and also for not warning you about the falling boulder; in this case I cause your death by pushing the boulder, and I also cause your death by failing to warn you about the falling boulder).

Now return to the example of the Queen and the Prince. Imagine, again, that the Queen asks the Prince to water her plant. Imagine that the Prince decides to spend the afternoon savoring some oaten biscuits on his lounge chair instead of watering the plant. Imagine, moreover, that his eating the biscuits doesn't *result in* his failing to water the plant

(as it does in the amnesia-inducing biscuit scenario described above): he simply eats the biscuits *instead of* watering the plant. Now consider two different scenarios of this sort:

1. Had the Prince not eaten the biscuits, he still wouldn't have watered the Queen's plant: he would have attended a theater performance instead.
2. Had the Prince not eaten the biscuits, he would have watered the Queen's plant.

Imagine, for example, that the Prince keeps a list of possible things to do at different times on a given day, ranked by order of preference. In the second scenario watering the Queen's plant is second on that list. So, had he not eaten the biscuits, he would have watered the plant. In the first scenario, by contrast, watering the plant is ranked lower, perhaps third, or fourth, or at the very bottom of the list. So, in that scenario, had he not eaten the biscuits, he still wouldn't have watered the plant.

Now, on the face of it, the causal relation between the Prince and the Queen's plant is of the same kind in both scenarios: it is purely *negative*. It seems clear that, in both cases, the Prince causes the plant to die by failing to water it, i.e. by failing to interfere with the natural processes that led to the plant's drying up and dying. And it seems clear that this is the only way in which he contributes to the plant's death: he contributes to the plant's death by failing to act in a certain way, and not by acting in a certain way. Or, at least, this is true of the Prince's contribution to the plant's death *at the relevant time, on that particular afternoon*: nothing he did *then* contributed to the plant's death.⁶

However, CC entails otherwise. Consider the counterfactual:

⁶ Maybe something he did *earlier* contributed to the plant's death. This would be the case if, e.g., he had poisoned the Queen's plant earlier that day and, as a result, the plant needs to be watered now in order to survive.

Had the Prince not eaten the biscuits, the plant's death wouldn't have occurred.

This counterfactual is false in the first scenario, but true in the second scenario. Moreover, the counterfactual dependence that exists between the Prince's eating the biscuits and the plant's death in the second scenario is an ordinary dependence between fully distinct events (it is not backtracking, logical, mereological, etc.).⁷ So CC entails that the Prince's eating the biscuits is a cause of the plant's death in the first scenario (whereas it doesn't entail this about the first scenario). Thus, according to CC, whereas the Prince's contribution to the plant's death (at the relevant time) is purely negative in the first scenario, it is *mixed* in the second: in the second scenario, CC entails that the Prince caused the plant's death by failing to water the plant, and also by eating the biscuits. This is the wrong result; again, the Prince's contribution to the plant's death (at the relevant time) is of the same type in both cases: it is a purely negative contribution.

As I explained above, it is possible to imagine a scenario of a different kind where the Prince's eating the biscuits does cause the plant's death. The scenario where the biscuits cause the Prince to forget to water the plant is a scenario of that kind. But, in the more ordinary type

⁷ Why isn't it backtracking? Someone could argue that it is backtracking by pointing out that, by the time the Prince started eating the biscuits, he had already made up his mind to not water the plant. So the counterfactual "Had the Prince not eaten the biscuits, the plant wouldn't have died" comes out true only if we assume that the Prince's antecedent process of deliberation resulted in a different outcome, in other words, only if we assume that the past is different in a certain way. However, this reasoning is not really backtracking, at least not in Lewis' sense. According to Lewis, we must allow for a "transition period" and imagine that the *immediate* past is different in some (non-fully specified) ways; otherwise there would be abrupt discontinuities in the facts and thus major violations of the laws (Lewis 1986c, pp. 39-40). So, unless the Prince made up his mind well in advance, the counterfactual dependence between the Prince's eating the biscuits and the plant's death would be of the ordinary type. Lewis' view aside, note that, if the counterfactual about the Prince is backtracking, so is *any* counterfactual stating that an agent would have intentionally done Y if he hadn't intentionally done X, when the agent is deciding between X and Y. But we assert counterfactual claims of this kind all the time.

of scenario I'm imagining, his eating the biscuits doesn't cause his failure to water the plant. And so neither does it cause the plant's death.

The problem generalizes. Many times we make "simple choices": we choose between doing one thing and doing some alternative thing. When we choose in favor of, say, A and against B, and then act accordingly, there's something we do and something we omit to do: we do A and we omit to do B. Accordingly, there are certain upshots that we cause in virtue of doing A and certain upshots that we cause in virtue of omitting to do B. Although there can be overlap between these two sets of effects, the overlap is not perfect, and it is usually only very minimal. In particular, not every upshot that we cause in virtue of omitting to do B is an upshot that we cause in virtue of doing A; in fact, very few upshots that we cause in virtue of omitting to do B are upshots that we also cause in virtue of doing A.⁸ Now, in circumstances of this type, where we would have done B if we had failed to do A, it is likely that, if an upshot U counterfactually depends on our *omitting to do B*, U will also counterfactually depend on our *doing A*. As a result, if we embrace CC, we end up with numerous unwanted positive causes. And thus we are committed to saying that, when an agent makes a simple choice and his contribution (at the relevant time) is purely negative, it is actually mixed.

This is bad because it clashes with the natural classification of types of causal contribution that we introduced above. But, more importantly for our purposes, it is bad because it conflicts with the initial motivation for accepting omissions as causes: *the inadequacy fact about positive causes*. We wanted to say that omissions can be causes partly because the positive events in the offing don't seem to do the job themselves. In particular, it seemed wrong to say that, if the Prince just ate some biscuits instead of watering the plant,

⁸ I draw attention to this fact in a different context in Sartorio (forthcoming).

then something he *did* that afternoon (eating the biscuits) caused the plant's death. We want to be able to say, instead, that the Prince's failure to water the plant caused the plant's death. By appeal to CC, we can say that. But the problem is that we *also* have to say (in simple choice scenarios) that something the Prince did (his eating the biscuits) caused the plant's death. So the counterfactual strategy truly backfires, for it lets in the wanted negative causes at the price of also letting in the inadequate positive causes, whose inadequacy was the reason we wanted the negative causes in the first place.

3. The problem of unwanted negative causes

In the preceding section we saw that CC is subject to the following objection: sometimes an upshot is caused by what an agent fails to do and not by what he does, but the upshot counterfactually depends both on what he fails to do and on what he does. In this section we will see that CC is also subject to the opposite objection: sometimes an upshot is caused by what the agent does and not by what he fails to do, but the upshot counterfactually depends both on what he does and on what he fails to do. So, again, CC entails that the agent's contribution to the outcome (at the relevant time) is mixed, when it isn't. As we will see, if the problem of unwanted positive causes was quite common, the opposite problem—the problem of unwanted negative causes—is even *more* widespread.

Imagine, again, that the Prince decided to not water the plant and to eat some oaten biscuits on his lounge chair instead. Imagine, also, that the high content of fat and sugar in the biscuits made him feel sick to his stomach. Finally, imagine that, had he not eaten the biscuits, he wouldn't have had a stomachache (for example, nothing else he ate that day would have made him sick), so the stomachache counterfactually depends on his eating the biscuits. Since

the counterfactual dependence in question is of the right kind (not backtracking, logical, mereological, etc.), CC entails that his eating the biscuits caused his stomachache. This is the right result. However, CC also entails that *the Prince's failure to water the plant* caused the stomachache. For, had the Prince watered the plant, he also wouldn't have had a stomachache (since he wouldn't have eaten the biscuits); and, again, the counterfactual dependence in question is not backtracking, logical, mereological, etc. But the Prince's failure to water the plant didn't cause his stomachache. If the Prince's contribution to the outcome of the plant's death (at the relevant time) was purely negative, his contribution to the outcome of his own stomachache (at the relevant time) was mainly *positive*: he caused it mainly by doing something (eating the biscuits) that resulted in his having a stomachache. I say "mainly" instead of "only" because, although the more salient cause of his stomachache was his eating the biscuits, there are other less salient causes of his stomachache, including, presumably, things he failed to do that afternoon, such as his not taking a powerful drug that would have prevented the stomachache.⁹ But, even if the stomachache had some (non-salient) negative causes consisting in the Prince's failure to do certain things that afternoon, it seems clear that his failure to water the plant *isn't* one of them. If the Prince's contribution to his own stomachache is mixed, it is not mixed by virtue of including the failure to water the plant as a negative cause.

A special kind of scenario in which it would be right to say that the Prince's failure to water the plant caused his stomachache is one where the Prince's failure to water the plant resulted in his eating the biscuits. For, if his failure to water the plant resulted in his eating the

⁹ Even if it is unreasonable to expect that the Prince would do any of those things (as we have seen, it is unreasonable to expect that the Queen would water my plant, but this doesn't prevent her from being a cause of its death in virtue of not watering it).

biscuits, then it would be plausible to say that the Prince's stomachache was caused (albeit indirectly) by his failure to water the plant. Imagine, for example, that his not watering the plant made him feel guilty, which triggered his hunger, which caused him to eat the biscuits. In that case his failure to water the plant would also be a cause of the stomachache. But, in the original case, where he simply eats the biscuits *instead of* watering the plant, not *because* he didn't water the plant, his failing to water the plant is not similarly a cause of the stomachache.

The problem of unwanted negative causes generalizes to many omissions by the Prince at the time, namely, the omissions of (most of) those acts that eating the biscuits on his lounge chair *precludes*: his not attending a theater performance, his not riding a horse, his not going on a hunting trip, etc. CC is likely to entail that each of these omissions also caused his stomachache because, if the Prince had engaged in any of those activities, he wouldn't have had a stomachache (because he wouldn't have eaten the biscuits). In other words, CC gets the wrong result for (most) omissions of acts that are incompatible (physically incompatible, or incompatible as a matter of physical necessity) with the act that the Prince actually performed and that caused his stomachache.¹⁰ Thus, besides entailing that the Prince's contribution to the stomachache is mixed in a certain way, CC entails that there exist, in addition to the relevant positive causal relation, *countless* negative causal relations that aren't really there.

In this respect, the problem of unwanted negative causes is more general and widespread than the problem of unwanted positive causes from the previous section. As we have seen, the problem of unwanted positive causes is the problem that arises when there is counterfactual dependence between actions and the upshots of omissions. In contrast, the

¹⁰ Most of them, but not *all* of them, because eating a dozen burgers would also have resulted in a stomachache.

problem of unwanted negative causes is the problem that arises when there is counterfactual dependence between omissions and the upshots of actions. But the number of things we fail to do at any given time is usually, if not always, much larger than the number of things we do at that time. As a result, there are many more “fake” counterfactual dependencies between *omissions* and upshots than between *actions* and upshots. In other words, there are many more unwanted *negative* causes than unwanted *positive* causes.

There is a second respect in which the problem of unwanted negative causes is more general than the problem of unwanted positive causes. As we have seen, the problem of unwanted positive causes arises when certain dependencies obtain “accidentally.” It is, in a sense, an accident that the Prince was deciding between only two alternatives, and thus, that, if he hadn’t eaten the biscuits, he would have watered the plant. As a result, this problem doesn’t generalize to all acts and omissions of physically incompatible acts, only to those where an agent makes what I have called a “simple choice.” By contrast, the problem of unwanted negative causes arises for *any* act and *any* omission of a physically incompatible act. For, arguably, it is always true that, when doing A and doing B physically exclude each other, if the agent had done B, he wouldn’t have done A.

I conclude that, just as there is a problem of unwanted positive causes for counterfactual theories of causation, there is also a problem of unwanted negative causes. In addition, the latter problem is more general than the former problem, in two respects. First, there are *always* unwanted negative causes (not so for unwanted positive causes: there are unwanted positive causes only sometimes). And, second, there are always *several* unwanted negative causes (not so for positive causes: even when there are some unwanted positive causes, there aren’t always several of them).

The problem of unwanted negative causes is a serious problem because it shows that counterfactual views commit us to many more negative causes than we want to be committed to. But, more importantly for our purposes, it is a serious problem because it shows that, although counterfactual views can accommodate causation by omission, they do this by failing to respect the motivation for accepting that kind of causation: the inadequacy fact about positive causes. We wanted to say that omissions are causes when positive causes were inadequate to account for the occurrence of certain outcomes. But we *don't* want to say that omissions are causes when the positive causes are perfectly adequate to account for those outcomes! However, this is what we would have to say, if we were to endorse CC.

4. Commensuration

Briefly, the Prince of Wales problem for counterfactual theories of causation (and, in particular, for CC) is this. As we have seen, there seems to be an important distinction between the outcomes that we cause by *doing* certain things and the outcomes that we cause by *failing* to do certain things. More generally, there seems to be an important distinction between what results from how things *are* and what results from how things are *not*; these sets of outcomes are usually not the same. However, there can easily be counterfactual dependence between how things *are* and the outcomes of how things are *not*; conversely, there can easily be counterfactual dependence between how things are *not* and the outcomes of how things *are*. Hence counterfactual dependencies can, quite ordinarily, be “fake” in that they can fail to reveal genuinely causal connections in the world.

The Prince of Wales problem is particularly puzzling because it initially seemed as if a counterfactual theory, armed with CC, was in an optimal position to account for the

possibility of causation by omission. For, on the face of it, outcomes can counterfactually depend on omissions and other absences just like they can counterfactually depend on (positive) events, and it is quite plausible to think that counterfactual dependence is all it takes for causation by omission to take place. Now, this initial advantage of counterfactual theories appears to be lost (or at least significantly reduced) once we realize that CC commits us to numerous unwanted causes, both of the positive and the negative kind, including many of those unwanted positive causes whose “inadequacy” motivated the search for negative causes in the first place and, also, many unwanted negative causes operating in those cases where the positive causes seemed perfectly adequate.

How could a counterfactual theorist try to address the Prince of Wales problem? In this section I discuss one main attempt to address it. The proposed solution I have in mind consists in abandoning CC, while at the same time trying to preserve the “spirit” of a counterfactual view.

One natural way to try to do this would be to appeal to Yablo’s idea that causes are “commensurate with” or “proportionate to” their effects, where the relevant notion of commensurability or proportionality is itself spelled out in counterfactual terms (Yablo 1992). The proposal is complicated, but the rough idea is that a cause is something that has just the right kind of “essence” for its effect: nothing with a poorer essence would have been sufficient for the effect to happen, and nothing with a richer essence was necessary for the effect to happen. For example, imagine that, when Socrates drank the hemlock, he guzzled it. Arguably, Socrates’ *guzzling* the hemlock didn’t cause his death because he would still have died if he hadn’t guzzled it, but if he had drunk it more slowly. Something with a poorer essence than his guzzling the hemlock (his simply drinking it) would have been sufficient for

his death; thus the guzzling isn't a cause of the death. This is so, Yablo says, even if, as a matter of fact, Socrates wouldn't have drunk the hemlock without guzzling it (for example, if he was a sloppy eater); even in this case, it was the drinking and not the guzzling that caused the death. Also, let's imagine that a bridge collapsed after one of its bolts suddenly snapped. The bridge is built in such a way that, given time to respond, it can shift its weight away from any failing bolts; so, if the bolt had snapped less abruptly, the bridge wouldn't have collapsed. Arguably, the bolt's simply snapping didn't cause the bridge's collapse; its *suddenly* snapping did. This is so because something with a richer essence than its simply snapping was needed for the bridge to collapse. And, again, Yablo claims, we would still want to say that the sudden snapping, not the snapping, caused the collapse, even if the circumstances are such that the bolt wouldn't have snapped at all if it hadn't done so suddenly (for example, if the temperature was low enough that bolts snap suddenly if at all). So, Yablo concludes, causes are commensurate with their effects in that they have the right amount of specificity built into their essences: nothing less specific would have done and nothing more specific was needed.

Note that, if Yablo is right, there are reasons to reject CC that are independent from those discussed in this paper (in the sense that they are provided by positive causes alone). For, on Yablo's view, Socrates' guzzling the hemlock wouldn't cause Socrates' death even if the death counterfactually depended on the guzzling (e.g., if Socrates was a sloppy eater and thus he wouldn't have drunk the hemlock unless he guzzled it). Also, on Yablo's view, the bolt's simply snapping wouldn't cause the bridge's collapse even if the collapse counterfactually depended on the snapping (e.g., if the bolt wouldn't have snapped at all unless it snapped suddenly). Of course, all of this rests on a fine-grained conception of events according to which different events can occupy the same spatio-temporal location, as long as

they differ in their modal properties, which is something that friends of more coarse-grained views of events would not accept.¹¹ But I am not interested in assessing the prospects of the commensuration view as an objection to CC here. Instead, I'm interested in examining the prospects of that view as a way of *improving on CC*.

How could the commensuration view be used to improve on CC? Take, first, the fact that the Prince's eating the biscuits didn't cause the Queen's plant's death. On the basis of the commensuration view, we could say that this is because something with a "poorer essence", namely, the Prince's failure to water the plant, would have been sufficient for the plant's death. (The Prince's eating the biscuits has a "richer essence" than his failing to water the plant in the sense that it takes more for him to eat the biscuits than it does for him to fail to water the plant: if he was eating the biscuits on his lounge chair, he was at the very least not watering the plant, but there are other requirements on what he was doing.) So his eating the biscuits isn't commensurate with the plant's death, even if the plant's death counterfactually depends on his eating the biscuits. Also, we could say that the Prince's failure to water the plant didn't cause his stomachache because something with a richer essence, namely, his eating the biscuits, was needed for his stomachache to obtain. Again, his failure to water the plant isn't commensurate with the stomachache, even if the stomachache counterfactually depends on his failure to water the plant.

Importantly, following Yablo, one could try to cash out all of these claims in purely counterfactual terms, thus preserving the spirit of a counterfactual view. For example, we could say that the eating of the biscuits didn't cause the plant's death because something less

¹¹ E.g., Davidson 1967. The distinction between actions and omissions on which this paper rests is, I think, much more basic and established than the distinction between, say, the event of Socrates' guzzling the hemlock and the event of his simply drinking it.

specific than the eating of the biscuits was sufficient, and that this is true in virtue of the truth of the following counterfactual:

If the Prince's failure to water the plant had obtained *without his eating the biscuits*, then the plant's death would still have occurred.

Also, we could say that the failure to water the plant didn't cause the stomachache because something more specific was needed, and that this is true in virtue of the truth of the following counterfactual:

If the Prince's failure to water the plant had obtained *without his eating the biscuits*, then the stomachache wouldn't have occurred.

With this in mind, we could revise CC as follows:

(CC*) If there is (ordinary) counterfactual dependence between C and E (where C and E are fully distinct), *and if nothing less specific than C is sufficient for E and nothing more specific than C is needed for E*, then C is a cause of E.

where the locutions "less/more specific", "sufficient", and "needed" have to be understood as explained above. To be clear: CC* is not Yablo's own preferred approach. Whereas CC* states a sufficient condition on causation, Yablo regarded commensuration more as a

necessary condition on causation than as a sufficient condition.¹² Also, he never tried to apply the proposal to negative causes. Still, I think that Yablo's ideas about commensuration can be naturally applied to our problem at hand, and this is what CC* attempts to capture.

Now, is retreating to CC* a satisfactory solution to the Prince of Wales problem? I will offer two main reasons to be skeptical.

First, it's important to realize that, on this picture, counterfactual dependence is no longer sufficient for causation (if anything, it's *commensuration* that's sufficient for causation). This can still be within the spirit of a counterfactual view if, as Yablo suggests, commensuration itself can be spelled out in counterfactual terms. But the resulting view no longer has the neatness and simplicity of the counterfactual criterion that might have made it seem attractive in the first place (or what was left of such neatness and simplicity after the necessary refinements discussed when we first introduced CC).

Second, retreating to CC* introduces new problems of its own for the idea that we can give sufficient conditions for causation in counterfactual terms. Interestingly, these are *preemption* problems.

Let me explain. Consider, again, the relation between the Prince's failure to water the plant and his stomachache. Let 'C' be the non-watering and 'E' the stomachache. E counterfactually depends on C: had the Prince watered the plant, he wouldn't have had a stomachache (because he wouldn't have eaten the biscuits). CC* attempts to avoid the implication that the non-watering caused the stomachache by claiming that the non-watering isn't commensurate with the stomachache. The non-watering isn't commensurate with the

¹² I think this is because he took the consensus to be that the problems for the sufficiency of counterfactual dependence for causation are more serious than those for its necessity. This might be right of the time when he wrote his article. As I have pointed out, the consensus now appears to be that the opposite is true.

stomachache because something more specific (the eating of the biscuits; call it 'D') was needed for the stomachache to occur. And this is supposed to be captured by the fact that, had the Prince not watered the plant *while failing to eat the biscuits*, the stomachache wouldn't have occurred (had C occurred without D, E would not have occurred). But we may imagine a slightly different scenario where, had C occurred without D, E would still have occurred. For example, imagine that, had the Prince failed to water the plant and failed to eat the biscuits, he would have gone for a walk outside, when it was bitterly cold, and this would equally have resulted in a stomachache. In that scenario D wasn't needed for E to occur (although of course it still causes E). So, presumably, in that case CC* would entail that the non-watering causes the stomachache.¹³ But, again, this is the wrong result.

In other words, once we move from CC to CC*, the counterfactual criterion faces problems that used to be just problems for the claim that counterfactual dependence is necessary for causation. Preemption becomes a problem, not just for the idea that counterfactual dependence (or something close to it) is necessary for causation, but also for the idea that counterfactual dependence (or something close to it) is sufficient for causation. This is not a promising result.

5. Conclusion

I have argued that counterfactual views of causation cannot accommodate causation by omission while remaining faithful to the motivation for accepting that kind of causation. In this final section I attempt to generalize this result.

¹³ CC* would only entail this if nothing more specific was needed (and if no event with a poorer essence was sufficient). Of course, I haven't shown this. But I take it that Yablo's idea is that the event that plays the role of D is the event that is actually the cause. And that event is, in this case too, the eating of the biscuits.

Some philosophers reject the possibility of causation by omission. Still, some of these philosophers feel the need to provide an account of the relation that omissions can bear to outcomes, in virtue of which agents can be morally responsible for those outcomes. Notably, Dowe has offered an account of a relation of this kind, which he called “quasi-causation” (Dowe 2000, 2001, and 2004). According to Dowe, omissions can quasi-cause outcomes, and it is in virtue of this relation of quasi-causation that agents can be morally responsible by omission. Quasi-causation is a counterfactual relation. Basically, whereas for a counterfactual theorist of causation who believes in causation by omission counterfactual dependence between an omission and an outcome is the mark of a causal relation between the omission and the outcome, for Dowe it is the mark of a quasi-causal relation. As far as I can see, everything I’ve said about the prospects of a counterfactual theory of causation can be said, *mutatis mutandis*, about a counterfactual theory of quasi-causation like Dowe’s.

Briefly, for someone like Dowe, the problem arises as follows. We think that agents can be responsible for outcomes in the world in different ways. Sometimes they are responsible in virtue of having caused those outcomes. Other times they are responsible without causing those outcomes. So we need to find a new way to account for the agents’ responsibility in these scenarios. Let’s call this fact “the inadequacy fact about *causes*.” The inadequacy fact about causes motivates the search for a new theory, a theory of “quasi-causation.” It is natural to try to give such a theory in counterfactual terms. But a counterfactual theory of quasi-causation would face the Prince of Wales problem. For similar arguments to those offered here would show that there are “fake” counterfactual dependencies between genuine causes and the upshots of quasi-causes as well as between quasi-causes and the upshots of genuine causes. As a result, a counterfactual theory of quasi-causation would

fail to respect the initial motivation for giving a theory of that kind, i.e. the inadequacy fact about causes.

To conclude: the Prince of Wales problem is not just a problem for counterfactual theories of causation. It is a more general problem that arises for any theory that attempts to understand the contribution of omissions in counterfactual terms.

References

- Beebe, H. (2004) "Causing and Nothingness," in J. Collins, N. Hall, and L.A. Paul, eds., *Causation and Counterfactuals*, M.I.T. Press, 291-308.
- Davidson, D. (1967) "Causal Relations," *Journal of Philosophy* 64.
- Dowe, P. (2000) *Physical Causation*, Cambridge U.P.
- Dowe, P. (2001) "A Counterfactual Theory of Prevention and 'Causation' by Omission," *Australasian Journal of Philosophy* 79, 2, 216-26.
- Dowe, P. (2004) "Causes are Physically Connected to their Effects: Why Preventers and Omissions are not Causes," in C. Hitchcock (ed.) *Contemporary Debates in Philosophy of Science*, Blackwell, 2004, chapter 9.
- Hall, N. (2004) "Two Concepts of Causation," in J. Collins, N. Hall, and L.A. Paul, eds., *Causation and Counterfactuals*, M.I.T. Press, 225-76.
- Kim, J. (1973) "Causes and Counterfactuals," *The Journal of Philosophy* 70, 17: 570-2.
- Lewis, D. (1986a) "Causation," in *Philosophical Papers*, vol. II, New York: Oxford U.P., 159-213.
- Lewis, David (1986b) "Events," in *Philosophical Papers II*, Oxford U. P., pp. 241-269.

- Lewis, D. (1986c) "Counterfactual Dependence and Time's Arrow," in *Philosophical Papers*, vol. II, New York: Oxford U.P., 32-66.
- Lewis, D. (2002) "Causation as Influence," reprinted in J. Collins, N. Hall, and L.A. Paul, eds., *Causation and Counterfactuals*, M.I.T. Press, 2004: 75-106.
- Lewis, D. (2004) "Void and Object," in J. Collins, N. Hall, and L.A. Paul, eds., *Causation and Counterfactuals*, M.I.T. Press: 277-90.
- McGrath, S. (2005) "Causation by Omission: A Dilemma," *Philosophical Studies* 123: 125-48.
- Sartorio, C. (2004) "How to Be Responsible for Something Without Causing It," *Philosophical Perspectives* 18: 315-36.
- Sartorio, C. (forthcoming) "Omissions and Causalism," *Noûs*.
- Schaffer, J. (2000) "Causation by Disconnection," *Philosophy of Science* 67: 285-300.
- Schaffer, J. (2004) "Causes need not be Physically Connected to their Effects: The Case for Negative Causation," in C. Hitchcock (ed.) *Contemporary Debates in Philosophy of Science*, Blackwell, 2004, chapter 10.
- Thomson, J. (2003) "Causation: Omissions," *Philosophy and Phenomenological Research* 66, 1: 81-103.
- Yablo, S. (1992) "Cause and Essence," *Synthese* 93: 403-49.