

Chapter 26: Causation And Ethics*

In this article I examine potential applications of the concept of cause to some central ethical concepts, views, and problems. In particular, I discuss the role of causation in the family of views known as consequentialism, the distinction between killing and letting die, the doctrine of double effect, and the concept of moral responsibility.

My main aim is to examine the extent to which an appeal to the concept of cause contributes to elucidating moral notions or to increasing the plausibility of moral views. Something that makes this task interestingly complex is the fact that the notion of causation itself is controversial and difficult to pin down. As a result, in some cases the success of its use in moral theory hinges on how certain debates about causation are resolved. I will point to examples of this phenomenon as I proceed.

1. Causation And Consequentialism

Consequentialist theories are theories according to which the moral status of an act is exhausted by the moral status of its consequences.¹ What are the consequences of an act? Numerous answers are possible. We could take them to be its causal consequences. Or we could

take them to be its causal and logical consequences. Or we could take them to be the whole state of the world following the act. And so on. Different answers generate different versions of consequentialism.² Here I will discuss the causal answer in particular. I will dub it ‘causal consequentialism.’

The content and implications of causal consequentialism depend on the nature of causation. The concept of cause appears to be restrictive in some ways and inclusive in others. It is restrictive in that consequences of an act that are too intimately related to it are not causal. For instance, an act of killing can be said to have the death of the victim as a consequence, but the relation between the two is presumably not causal.³ For the fact that there was a killing *entails* the fact that there was a death, and an entailment is not a causal relation (consider: the fact that I entered the room at noon entails the fact that someone entered the room at noon, but it didn’t cause it). Thus, according to causal consequentialism, the victim’s death would not be a consequence of the killing, and thus it would not contribute to the moral status of the killing. On the other hand, the concept of cause also appears to be quite liberal in that every act has many effects extending indefinitely into the future. Causal chains can be very complex and they can link events in unexpected ways.

These features of the concept of cause have interesting implications for causal consequentialism. I will examine some of them.

First, the restrictive side of causation has the following implication. Imagine that an assassin murders a child's parents. When he does, the child becomes an orphan. But the assassin's murdering the parents doesn't *cause* the child's orphanhood, for the fact that the child's parents were murdered *entails* the fact that the child is an orphan, and entailment is not causation. Thus causal consequentialism would imply that the child's orphanhood is not among the consequences of the act of murdering of the parents. This seems wrong because we want to blame the assassin for murdering the parents and making the child an orphan as a result. On the basis of a similar example, David Sosa has argued that a consequentialist should embrace a broader concept of consequence (Sosa (1993)).

Now, the causal consequentialist could reply that, although the assassin's murdering the parents didn't cause the child's orphanhood, some other act by the assassin did, e.g., his pulling the trigger (note that the fact that he pulled the trigger doesn't entail the fact that the child became an orphan; after all, he could have missed).⁴ Thus we can still blame the assassin for the child's orphanhood on these grounds. However, an important part of the objection survives. Causal consequentialism would still not entail that the child's orphanhood accounts for the moral status of the act of murdering the parents, for *that* act didn't cause the child's orphanhood. And it is plausible to suggest that the immorality of that act is explained, at least in part, by the fact that it has the child's orphanhood as a consequence. (Or, at least, it makes sense to suggest that a

consequentialist would want to say this.) In addition, there might be cases where, although an agent is to blame for an outcome, *no* act by the agent can be said to be a cause of the outcome. I will discuss such cases in section 4.

We have examined reasons to believe that the concept of cause might be too narrow for consequentialism to draw on it. Now let us examine if there is also a sense in which it might be too broad.

A common objection (or family of objections) to consequentialism is based on the observation that any act has too many consequences, more than could plausibly matter to the moral status of the act. Call this objection ‘the irrelevant consequences objection.’ One version of this objection points to the fact that many consequences of acts are unforeseen (if you look far enough down the chain of events), and thus consequentialism doesn’t provide a criterion that one can easily appeal to in deciding what to do (Mill anticipates this kind of objection to his view in his (1863)). Imagine, for instance, that an apparently morally good act indirectly influenced the time at which Hitler’s parents conceived a child. It is initially implausible to suggest that the fact that the act led eventually to genocidal consequences makes the act wrong. There are two standard consequentialist responses to this version of the objection. First, the consequentialist could reply that we shouldn’t expect the theory to play the role of a decision procedure; after all, consequentialism is a principle about the morality of acts, not about how we can come to know

what the right moral act is. Therefore the act might be immoral, though we couldn't have known it at the time (see, e.g., Bales (1971)). Second, a consequentialist might want to retreat to a version of consequentialism according to which the moral status of an act is determined, not by the whole set of consequences of the act, but only by its "expected" consequences, i.e. the consequences that the agent could reasonably expect to occur (see, e.g., Jackson (1991)).

Still, there is a version of the objection that survives both of these moves. Imagine that I can choose between pulling the trigger of a gun and not doing so. If I do, a person will die; if I don't, then a hit man who is waiting in the background will shoot and the victim will still die. (This type of case is called a "preemption" case in the causation literature.) Intuitively, under these circumstances, I ought not shoot. But consequentialism seems to entail that I should be indifferent, given that the consequences of my shooting and my not shooting are the same.⁵ In this case, it is not the lack of foreseeability, or the remoteness, of the consequences of not shooting that makes them intuitively irrelevant, but it is the idea that they are not really "due to the agent." Thus the objection cannot be addressed by, e.g., retreating to an expected consequences version of consequentialism.

How does causal consequentialism fare with respect to this version of the irrelevant consequences objection? At first sight, it might seem that it is not particularly well placed to answer it. For, as I have pointed out, according to causal consequentialism, *any* item in the causal

chain flowing from an act would seem to be a consequence of it. However, there are at least two ways in which a causal consequentialist could try to resist this implication. I will examine the prospects of each of these attempts.

First, the causal consequentialist might want to rely on the notion of “proximate causation” discussed in the philosophy of law. (Sinnott-Armstrong points to this possibility in his (2003) but says that it hasn’t been developed as a version of consequentialism.) On this view, an agent is a cause of an outcome only if the outcome “flows from his agency”: the interventions of other agents or of unexpected natural phenomena can break up causal chains and rid an agent of causal responsibility for an outcome [REF 38]. This view entails that certain unexpected outcomes and other remote consequences of an act are not its genuine consequences. Also, it entails that the victim’s death is not a consequence of my failure to shoot in the hit man case, given that it occurs via the intervention of another moral agent. Thus this view escapes the irrelevant consequences objection, even on its more recalcitrant version. But the price paid is too high. Given the role that factors like expectations, norms, agency, etc. play in this view, it is hard to square it with the idea that causation is an objective, natural, and “out in the world” relation. Thus I suggest that we set this view aside and look for a more promising alternative.⁶

Fortunately for the causal consequentialist, there is a better approach: to reject the transitivity of causation. Some philosophers have rejected the transitivity of causation (see, e.g.,

Hitchcock (2001) and Yablo (2002)) but, as far as I know, no one has explicitly drawn on it to build a more plausible version of consequentialism. Rejecting the transitivity of causation could help the causal consequentialist in the following way: if causation isn't transitive, then the existence of long causal chains linking events doesn't imply that the events are causally related. On these grounds, a causal consequentialist could claim that not every event in the chain flowing from a given act contributes to the moral status of the act.

To be clear: rejecting the transitivity of causation doesn't amount to claiming that, when *c* causes *d* and *d* causes *e*, *c* never causes *e*, but only that it doesn't *sometimes* (and that whether or not it does depends on the specific features of each case). However, rejecting transitivity can help the causal consequentialist answer the objection posed by the hit man case in particular. For the hit man case seems to be a scenario where transitivity fails, if it fails at all. Let me explain.

Someone who rejected the transitivity of causation could say the following about the hit man case. If I refrain from shooting, my failing to shoot causes the hit man to shoot, and the hit man's shooting, in turn, causes the victim to die; however, my failing to shoot doesn't cause the victim to die. For this is a case where transitivity fails: intuitively, I am only a cause of the person's death if I shoot, not if I don't. This is an independently plausible claim. It is also consistent with an objective, natural, and "out in the world" conception of causation. In particular, it is not the existence of a second *agent* in the causal chain that makes us want to say

that my failing to shoot isn't a cause. To see this, note that we can construct cases with the same causal structure but deprived of moral agents. Here is an example: yesterday's rain caused the existence of a puddle today; however, if it hadn't rained, the ensuing dryness would have caused a bowl with water to crack, and the water would have filled the hole in a similar way. Intuitively, the rain caused the puddle, but the absence of rain wouldn't have caused the puddle - although it would have caused the crack in the bowl, and the crack in the bowl would have caused the puddle. The hit man case shares this structure. So it is plausible to claim that, if I don't shoot the victim, the hit man causes his death but I don't.⁷

I conclude that a causal consequentialist would benefit from a conception of causation that rejected transitivity. The hit man objection rests on the intuition that, although a killing would have occurred regardless of what I do, if I decide not to pull the trigger, I am not the one who "does it" and thus I am not responsible for the death. As we have seen, by rejecting the transitivity of causation the causal consequentialist can capture this intuition without stepping outside of the consequentialist framework. Note, however, that the ultimate success of such a project depends on whether causation does turn out to be a non-transitive relation, i.e. it depends on how a certain debate in the metaphysics of causation is resolved.⁸

In sum, causal consequentialism appears to have some advantages and some disadvantages. A potential disadvantage is that it ignores possible non-causal consequences of

acts, such as logical consequences, which might be relevant to the morality of acts. On the other hand, a potential advantage is that, armed with a non-transitive concept of cause, causal consequentialism can rebut some serious objections to consequentialism in an appealing way. Ideally, consequentialism would benefit from a concept of consequence that combined the benefits of the causal concept with those of a broader notion.

2. Causation And Killing/Letting Die

Some philosophers believe that killing is (other things being equal) morally worse than letting die (see, e.g. Foot ([1984] 1994) and Quinn ([1989*a*] 1994)). Call this thesis ‘KLD.’ If KLD were true, it could be used to explain some ordinary moral judgments. For instance, it could explain why some people think that failing to give to charity is not wrong (or at least less bad than, say, sending poisoned food to third-world countries), on the grounds that it is merely letting people die. Also, it could account for a distinction drawn in contemporary medical practice between “active” forms of euthanasia (such as injecting a terminal patient with a lethal dose of a drug), which are generally regarded as impermissible, and “passive” forms (such as withholding a certain medical treatment), which are generally regarded as permissible, on the grounds that active euthanasia is killing and passive euthanasia is letting die.

If there were a moral difference between killing and letting die, what could account for it? Here I will look at attempts to account for it in causal terms.

The simplest way to build a causal account of KLD is to start by grounding the killing/letting die distinction on the action/omission distinction, and to claim that the action/omission distinction coincides with the distinction between causes and non-causes. Roughly, the proposal would be that a killing is the causing of a death because a killing involves an action and actions are causes, whereas a letting die is not the causing of a death because a letting die involves an omission and omissions are not causes (see, e.g., Callahan (1989)).

There are several problems with this type of approach. The first problem, which I'll have to bypass here, is the identification of the killing/letting die distinction with the action/omission distinction. As it has been pointed out, it seems possible to kill by omitting to do something (as when a kidnapper kills his victim by starving him, or by failing to give him the insulin he needs) and, conversely, it seems possible to let die by doing something (as when a respirator is turned off to let a patient die a peaceful death). (See, e.g., Quinn ([1989a] 1994).)

But, setting this objection aside, there are also problems with the claim that actions can be causes but omissions cannot. Whether omissions can be causes is a highly debated issue in the literature on causation [REF 19]. Again, then, the success of this proposal hinges on how a certain debate in the metaphysics of causation is resolved. More importantly, however, the

following dilemma arises for someone who wants to take this line. Common sense dictates that at least some omissions are causes. Thus, on the one hand, if common sense is right and omissions can be causes, the appeal to the action/omission distinction doesn't help account for KLD. On the other hand, suppose that common sense is wrong and omissions cannot be causes. Then, I take it, the reason the proposal might seem promising to anyone is that, intuitively, there is a clear moral distinction between causing deaths and not causing deaths. That is to say, the thought would have to be that, intuitively, causing harm is bad (at least under certain circumstances) and not causing harm isn't, so killing someone can be bad because it is the causing of a death but letting someone die isn't bad because it isn't the causing of a death. However, this is too strong: surely, one can merely let someone die and still act very badly. For instance, if I see a seriously injured person on my way home and I don't stop to save him to prevent my take-out pizza from getting cold, I don't kill the person (I merely let him die) but I still act wrongly. Consequently, it seems that, if one embraced the view that omissions cannot be causes, then one would have to acknowledge that not causing harm can be bad. As a result, the project of grounding KLD would still fail, for then the connection between not causing harm and acting permissibly (or less wrongly) would be severed. (For an argument along the lines of this second horn, see Howard Snyder (2002).)

In sum, the dilemma that a causal account of KLD would have to face is:

- (Premise) Either omissions can be causes or they cannot.
- (Horn 1) If omissions can be causes, then there is no causal difference between actions and omissions, and thus we cannot account for the killing/letting die distinction in causal terms.
- (Horn 2) If omissions cannot be causes, then we cannot ground the distinction between killing and letting die in the claim that causing harm is bad but not causing harm isn't, and thus we cannot account for the killing/letting die distinction in causal terms.
- (Conclusion) We cannot account for the killing/letting die distinction in causal terms.

Now, one could try to resist the dilemma by rejecting horn 1. There are two main ways of doing this. First, one could claim that some omissions can be causes, but not all of them can (in particular, not those involved in letting die cases). Second, one could argue that, even if all omissions can be causes, omissions have importantly different causal powers from those of actions (and those different causal powers generate the moral difference between killing and letting die). Note that these views would still be causal accounts of KLD. In what follows I briefly examine the prospects of each view.

Let us examine, first, the view that some omissions have causal powers but others don't. Presumably, the proposal would be that, e.g., when a kidnapper starves his victim to death, his failing to feed him causes the victim's death; by contrast, when a doctor fails to continue a medical treatment on a terminal patient, his discontinuing the treatment does not cause the patient's death (instead, the patient's ailment does). On this view, such causal difference explains why the doctor merely lets his patient die, and thus acts permissibly, whereas the kidnapper kills his victim, and thus acts impermissibly.

However, there is a serious objection to this view. What the view does, at bottom, is to attribute causal powers only to those omissions that are *salient* in the circumstances. The most salient threat to the patient is his ailment (not the doctor); the most salient threat to the kidnapped victim is the kidnapper. Thus, whereas the doctor's omission doesn't strike us as a cause, the kidnapper's omission does. However, salience is only a pragmatic factor. Pragmatic factors cannot make a metaphysical difference unless they are accompanied by other metaphysically relevant factors. Now, one could try to argue that in the doctor case there is a preexisting physical process (the disease) that, if left alone, will lead to the death, but in the kidnapper case there isn't such a process. If this were true, then there would be a metaphysical difference between the cases that could account for a causal difference. However, it is false: in the

kidnapper case too, there are some biological processes that will lead to the victim's death if he doesn't get the nutrients he needs to stay alive.

The difficulty encountered by this type of account of KLD is reminiscent of the "Queen of England problem" in the literature on causation by omission (see, e.g., Beebe (2004)). The Queen of England problem is the following. Imagine that we want to claim that the gardener's failure to water my plant was a cause of the plant's death. Then it seems that we'll have to acknowledge that everyone else's failure - including the Queen of England's - was a cause too. After all, there is no relevant metaphysical difference between the gardener's failure and the Queen of England's failure: the only ways in which they differ concern duties, expectations, etc. As a result, it seems that we should believe that the Queen of England causally contributed to my plant's death. The standard attitude in the literature (by those who accept the possibility of causation by omission) is to bite the bullet and accept this result (see, e.g., Lewis (2004)).

We have seen that the view that assigns different causal powers to different omissions faces serious difficulties. Now let us examine the view that actions (in general) and omissions (in general) have importantly different causal powers, the second suggested way of escaping the dilemma presented above. Roughly, this view claims that, although actions and omissions can both be causes, they cause things in importantly different ways, and this difference in causal powers is significant enough to generate a moral difference between killing and letting die.

What could be the difference in causal powers between actions and omissions?

Something that comes to mind is the distinction that some philosophers have drawn between “enabling” conditions (or “enablers”) and “triggering” conditions (or “triggers”) (see, e.g., Lombard (1990)). Roughly, an enabler is something that “facilitates” the occurrence of an effect, or merely makes it possible, without setting off the causal chain leading to it. Enablers are sometimes regarded as “background conditions”: facts or states of affairs that need to be in place for an outcome to happen, but that require a triggering event for the causal chain to unfold. On some views, enablers are not genuine causes (Lombard (1990) and also Thomson (2003)). But one could also argue that enablers *are* causes, although their contribution is different from that of triggers. For instance, one might want to distinguish the causal contribution of an enabler, like the oxygen present in the environment, without which a match wouldn’t have lit, from that of a trigger like the striking of the match. Similarly, one might want to distinguish the contribution of the mere existence of a bridge to a person’s act of crossing it from that of the reasons (beliefs and desires) for which he crossed it (I am adapting one of Thomson’s examples in her (2003)).

On the basis of the distinction between enablers and triggers, the next step would be to argue that omissions are always enablers, never triggers, and that this causal difference gives rise to a moral difference. For instance, it could be argued, the doctor’s failure to continue the treatment is only an enabler in that it only “facilitates” the occurrence of the patient’s death, with

the ailment being the trigger. Then the suggestion would be that, just like we normally regard the act of pulling a trigger as worse than the act of leaving a gun on a table unattended, similarly, we should regard an act of killing as worse than an act of letting die.

This view has two serious problems. First, at least sometimes, omissions - and absences in general - play the role of triggers. For instance, the kidnapper's failure to feed his victim "triggers" the victim's death, and the absence of rain "triggers" the drought. Thus the identification of omissions with enablers fails. Second, almost anything can be seen as a mere enabler. Take the striking of the match. Other things had to happen, besides the striking of the match, for it to light. So the striking of the match too, just like the presence of oxygen in the environment, is something that merely "facilitates" the occurrence of the effect, at least in the sense that it is not independently sufficient for the occurrence of the effect. Similarly, in the passive euthanasia case, it is confused to suggest that the ailment is independently sufficient for the death but the doctor's omission isn't: they need each other (as well as other conditions) to bring about the death. Again, it seems that the reason why we are tempted to draw a distinction between them is that one of them is more salient than the other, but salience isn't enough to ground a metaphysical difference. (Penelope Mackie raises a similar objection to Lombard's view of enablers in her (1992).)

I conclude that there are serious problems with the project of accounting for KLD in causal terms. The project faces a dilemma from which it is difficult to escape. Hence, if there is a moral difference between killing and letting die, it probably doesn't have a causal basis.

3. Causation And Double-Effect

On a classical formulation, the "doctrine of double effect" (DDE) says that an act that issues in a bad result is permissible if four conditions are met: first, the act is not bad in itself; second, the act also issues in a proportionally good result; third, the bad result is not intended but merely foreseen; fourth, the bad result is not a means to the good result (Mangan (1949)).

It is natural to understand the fourth condition - hereafter *the means condition* - causally: on this interpretation, the bad result is not a means to the good result when the bad result doesn't *cause* the good result. In turn, it is natural to interpret the third condition - hereafter *the intentionality condition* - as the subjective or psychological counterpart of the means condition. That is to say, if the means condition reads: 'The bad result is not a causal means to the good result', the intentionality condition reads: '*The agent* doesn't view the bad result as a causal means to the good result.' This is the version of DDE that I will be concerned with here. Some versions of DDE don't explicitly appeal to the concept of cause; instead, they make primitive use

of the means/end distinction or they analyze this distinction in other terms. I will argue that a *causal* version of DDE has an important advantage over other versions in that it can provide an appealing answer to a common objection to DDE.

Let us start by looking at some examples. DDE is generally used to account for the moral difference between cases like the following:

Terror Bomber: A pilot bombs civilians in order to lower the enemy's morale and thus end the war. The end of the war saves many lives.

Strategic Bomber: A pilot bombs a weapons factory in order to end the war. The bombing issues in the end of the war, but also in some civilian deaths. (Bratman (1987))

Intuitively, the terror bomber acts impermissibly but the strategic bomber acts permissibly.

Proponents of DDE claim that the difference consists in that, whereas the terror bomber intends the civilian deaths, or uses those deaths as means to his end, the strategic bomber doesn't: the civilian deaths are merely a foreseen side-effect of the strategic bomber's act and are not means to his end.

Also, DDE is often used to explain the difference between these two cases:

Transplant: A surgeon kills a healthy patient in order to use his organs to save five people. The one dies but the five live.

Trolley: A runaway trolley is hurtling down the tracks where five people are trapped. A bystander flips a switch that redirects the train towards a side track, where only one person is trapped. The one dies but the five live.

Intuitively, the surgeon in Transplant acts impermissibly but the bystander in Trolley acts permissibly. The problem of accounting for this difference has been called ‘the trolley problem (cf. Foot ([1967] 1994) and Thomson ([1976] 1986)). Proponents of DDE suggest the following solution to the problem: in Transplant, the surgeon intends the death of the one (without which he couldn’t save the five) and uses his death as a means to saving the five, but in Trolley the bystander doesn’t intend the death of the one and his death is not a means to the saving of the five (instead, it is a mere side-effect).

Now, a common objection to DDE is that the intended/foreseen distinction, as well as the distinction between one’s means and what results from one’s means, is fundamentally unclear (see, e.g., Davis (1984)). To illustrate the problem, consider Terror Bomber again. The advocate of DDE wants to claim that the terror bomber acts impermissibly because, unlike the strategic bomber, he uses the civilian deaths as means to his end and the deaths are an intended

consequence of his dropping the bombs. However, we are assuming that the terror bomber only bombs the civilians because he wants the war to come to an end: he doesn't do it just to kill some civilians. If there were a way for him to avoid their deaths and at the same time achieve his desired end (say, by misleading the enemy into thinking that the civilians are dead, thus still demoralizing them), then he would do it. This suggests that he doesn't really intend the civilians to die, but only to *seem* dead. Also, one could argue that in Terror Bomber the civilian deaths are not really a means to the war's coming to an end. The war would still have ended if the civilians had only seemed dead; hence, the means to the end of the war was only their seeming dead. Thus, the objection goes, DDE fails to explain the moral difference between Terror Bomber and Strategic Bomber.

The same objection could be raised for Transplant and Trolley. One could argue that, in Transplant, if the surgeon could somehow get the one patient's organs without killing him, he would do it. Hence, the surgeon doesn't really intend the one's death, but only the removal of his organs. Also, the means to the five's survival was not the one's death but the removal of the organs. Thus, the objection goes, DDE fails to explain the moral difference between Transplant and Trolley.

I think that the advocate of a *causal* version of DDE can answer this objection in a way that isn't available to other versions of the view. He could argue as follows. The truth of the counterfactual:

'Had the civilians seemed dead without being dead, the war would still have ended.'

doesn't entail that the civilian deaths didn't cause the end of the war. In the actual case, the deaths did cause the end of the war, *via* their causing it to be the case that the civilians seemed dead. Similarly, the truth of the counterfactual:

'Had the surgeon removed the organs from the one without killing him, the five would still have survived.'

doesn't entail that the one's death didn't cause the five's survival. In the actual case, the one's death did cause the five's survival, *via* its facilitating the removal of the organs from the one's body.⁹

In other words, the objection only works under the following assumption:

Counterfactual conception of means: If the counterfactual ‘Had *X* occurred without *Y*, *Z* would still have occurred’ is true, then *Y* is not a means to *Z*.

But, as it has been pointed out in the literature on causation, the concept of cause doesn’t seem to meet this simple counterfactual requirement. Consider a case of preemption: Fast Shooter shoots at Victim and Victim dies as a result; Slow Shooter also shoots but it takes longer for his bullet to arrive, so he is preempted by Fast Shooter. In this case, the counterfactual ‘Had Slow Shooter shot without Fast Shooter shooting, Victim’s death would still have occurred’ is true. However, Fast Shooter is still a cause of Victim’s death. This suggests that, if the notion of means is understood causally, as opposed to, e.g., merely counterfactually, the objection to DDE fails. (Again, this would depend, ultimately, on how the debate over the relation between causation and counterfactuals is resolved. But, at least initially, it seems that preemption cases undermine counterfactual theories of causation on their simplest versions [REF 8].)

I have formulated the reply to the objection in terms of the causal condition. But, arguably, it can be extended to the intentionality condition as well, which (as I am understanding it) is simply the psychological counterpart of the causal condition. Thus, an advocate of a causal version of DDE could claim that, in Terror Bomber, the pilot *does* intend the civilian deaths. To be sure, his goal *could* still have been achieved by the civilians only seeming dead. But in the

actual case it wasn't: in the actual case, the deaths caused the war to come to an end, and the terror bomber viewed those deaths as a causal means to his goal. Similarly, in Transplant, the surgeon *does* intend the one's death. Again, his goal *could* still have been achieved by the one's surviving the removal of his organs, but in the actual case it wasn't: in the actual case his goal was achieved by the one's death causing the saving of the five.

I have argued that a causal version of DDE has an important advantage over other versions of the doctrine. Namely, a causal interpretation of the notion of means, and of the related concept of an intended consequence, provides the doctrine with a promising answer to a standard objection.

4. Causation And Moral Responsibility

In this section I examine the relationship between causation and moral responsibility. I will not touch on issues that are the focus of another chapter, such as the relation between responsibility and causal determinism or the concept of "agent causation" [REF 25].

Causation seems to be related to moral responsibility in the following way. Agents are normally regarded as responsible for (some of) their actions and omissions but also for (some) external events and states of affairs. Which events and states of affairs? A natural suggestion is

that the *only* events and states of affairs that agents are responsible for are certain causal products of their actions and omissions. After all, the causal powers of an agent's actions and omissions seem to constitute the only links between the agent and the world in virtue of which the agent can impact the world. Call this 'the received view' about the relation between moral responsibility (for events and states of affairs in the world) and causation. According to the received view, again, an agent is morally responsible for something *only if* the agent caused it, i.e. only if some action or omission of his caused it. Plausibly, other conditions are required for the agent to be responsible, e.g., the agent could or should have foreseen that the outcome would result from his behavior, but the causal condition is usually taken to be a necessary condition. (For an example of a theory of responsibility that contains the causal condition as a component, see Feinberg (1970).) Here I will focus on the causal condition only. I will critically examine possible reasons to resist it.

How could an agent be responsible for something without causing it?¹⁰ Someone might think that the Sosa-type example from section 1 shows a way. Thus, one might think that an assassin can be responsible for a child's orphanhood in virtue of having killed the child's parents even though his killing the parents didn't *cause* the child's orphanhood. However, as I have argued in section 1, there are other actions by the assassin that arguably did cause the child's

orphanhood, e.g., his pulling the trigger of his gun. As a result, the causal condition is unscathed by this example: the assassin caused the child's orphanhood, since *some* act of his caused it.

I will argue, however, that there are other cases that threaten to undermine the causal condition. In what follows I put forth two such cases: one involving omissions and one involving commissions.

Case 1 (omissions): (For a more extensive discussion, see Sartorio (2004).) Imagine that two buttons have to be depressed at the same time to prevent a bad outcome from happening. Imagine that the agents in charge of the buttons independently fail to depress them because they want the harm to ensue. In that case, I submit, each agent is responsible for the outcome but neither agent caused it, in particular, neither of the individual omissions by the agents caused it. Briefly, the argument that the agents' omissions didn't cause the outcome goes as follows. Imagine that there had only been one agent: he was in charge of one button and an automated mechanism was in charge of the other button. In that case, if the mechanism failed, the agent's failure wouldn't be a cause (after all, there was nothing that the agent could have done to prevent the outcome, since the mechanism failed and both buttons needed to be depressed to prevent it). But whether the other button was being controlled by an automated mechanism or by another agent cannot possibly matter to the causal powers of the agent. So, in the case with two agents, neither agent's failure is a cause of the outcome. But, clearly, each agent bears some

responsibility for the outcome. Thus it is possible to be responsible for an outcome without causing it.

Case 2 (commissions): (For a more extensive discussion, see Sartorio (ms).) Imagine that a person is tied to a train track and that a runaway train is approaching him. There is a switch and a side track diverging from the main line, but the tracks reconverge before the spot where the person is located. Thus the train can kill the person via either route, assuming that the relevant pieces of track are in working order. Now imagine that there are two agents, Flipper and Reconnector. Flipper is by the switch and flips it when the train approaches it; as a result, the train runs on the side track for a while, then on the main track again, and ends up killing the person. Reconnector is by the segment of the main track where the tracks come apart. As it turns out, that piece of track had been disconnected earlier that day. As a result, if the train had run on the disconnected piece of track, it would have derailed and it wouldn't have killed the person. Reconnector reconnects that piece of track at around the same time that Flipper flips the switch.

I submit that this is another example of responsibility without causation. Note, first, that, had it not been for Flipper and Reconnector, the person wouldn't have died: the train would have continued on the main track and it would have derailed while passing through the disconnected segment. This is just like in Case 1, where the ensuing harm depends on the combined behavior of two moral agents. Thus, assuming that Flipper and Reconnector were fully aware of what they

were doing, it is likely that we want to blame them for the person's death. In particular, focus on Flipper's behavior: it seems clear that we want to blame him, at least partly, for the death.¹¹

However, I submit that his flipping the switch isn't a cause of the person's death. The argument for this parallels the earlier argument for Case 1. Imagine that what reconnected the segment of the main track had been an automated mechanism; in that case we wouldn't want to say that Flipper's redirecting the train was a cause of the person's death. After all, what he did was only to redirect a threat from one path to another, where, given the presence of the reconnecting mechanism, neither path was more threatening than the other.¹² But whether what reconnected the track was a mechanism or a moral agent cannot determine whether Flipper was a cause. So Flipper isn't a cause when what reconnects the track segment is Reconnector. Thus Flipper is responsible for the death without causing it.

I have argued that there is reason to believe that moral responsibility and causation are not related in the way suggested by the received view. That is to say, being morally responsible for an outcome doesn't require causing it. How are they related, then? I will offer the sketch of a suggestion. In Case 1, even if the two agents do not cause the harm, they are responsible for something that causes the harm. What causes the harm? The fact that the two buttons were not simultaneously depressed (since, had they *both* been depressed, the harm wouldn't have occurred). The two agents were responsible for this fact, and thus, given that *this* fact caused the

harm, they are responsible for the harm. Something similar is true of Case 2. Even though Flipper and Reconnector don't cause the harm, they are responsible for something that causes it. What causes the harm? The fact that the train ran on a viable track (since, had it not run on a viable track, it wouldn't have reached the person). Flipper and Reconnector are responsible for this fact and, given that *this* fact caused the harm, they are responsible for the harm. Now, this proposal gives rise to questions that I will have to set aside here, such as the threat of a possible regress (e.g., How are Flipper and Reconnector responsible for the further fact that the train ran on a viable track? Did they cause *that* fact?). My aim here has been only to make plausible the idea that the received view about the relation between responsibility and causation needs revision, and to offer a rough suggestion as to how to revise it. (For discussion of the alternative view, see Sartorio (2004).)

References

- Bales, R. E. (1971). 'Act-Utilitarianism: Account of Right-Making Characteristics or Decision Procedures?', *American Philosophical Quarterly*, 8: 257-65.
- Beebe, H. (2004). 'Causing and Nothingness', in J. Collins, N. Hall, and L. A. Paul (eds.), *Causation and Counterfactuals*. Cambridge, Mass: The MIT Press, 291-308.

Bratman, M. (1987). *Intentions, Plans, and Practical Reason*. Cambridge, Mass: Harvard University Press.

Callahan, D. (1989). 'Killing and Allowing to Die', *The Hastings Center Report*, 19: ??.

Davidson, D. ([1971] 1980). 'Agency', originally in R. Binkley, R. Bronaugh, and A. Marras (eds.), *Agent, Action, and Reason*, Toronto: University of Toronto Press, 1971. Reprinted in *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, 43-61.

Davis, N. (1984). 'The Doctrine of Double Effect: Problems of Interpretation', *Pacific Philosophical Quarterly*, 65: 107-23.

Feinberg, J. (1970). 'Sua Culpa', in *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton: Princeton University Press, 187-221.

Feldman, F. (1997). *Utilitarianism, Hedonism, and Desert*. New York: Cambridge University Press.

Foot, P. ([1967] 1994). 'The Problem of Abortion and the Doctrine of the Double Effect', originally in *Oxford Review*, 5, 1967. Reprinted in Steinbock and Norcross (1994), 266-79.

Foot, P. ([1984] 1994). 'Killing and Letting Die', originally in J. L. Garfield and P. Hennessey (eds.), *Abortion: Moral and Legal Perspectives*, Amherst: University of Massachusetts Press, 1984. Reprinted in Steinbock and Norcross (1994), 280-9.

- Hitchcock, C. (2001). 'The Intransitivity of Causation Revealed in Equations and Graphs', *Journal of Philosophy*, 98, 6: 273-99.
- Howard-Snyder, F. (2002). 'Doing vs. Allowing Harm', *The Stanford Encyclopedia of Philosophy (Summer 2002 Edition)*, Edward N. Zalta (ed.), =
<<http://plato.stanford.edu/archives/sum2002/entries/doing-allowing/>>.
- Jackson, F. (1991). 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection', *Ethics*, 101: 461-82.
- Lewis, D. (2004). 'Causation as Influence', in J. Collins, N. Hall, and L. A. Paul (eds.), *Causation and Counterfactuals*. Cambridge, Mass: The MIT Press, 75-106.
- Lewis, D. (1986). 'Events', in his *Philosophical Papers II*. Oxford: Oxford University Press, 241-69.
- Lombard, L. (1990). 'Causes, Enablers, and the Counterfactual Analysis', *Philosophical Studies*, 59: 195-211.
- Mackie, P. (1992). 'Causing, Delaying, and Hastening: Do Rains Cause Fires?', *Mind*, 101: 483-500.
- Mangan, J. (1949). 'An Historical Analysis of the Principle of Double Effect', *Theological Studies*, 10: 41-61.
- Mill, J. S. (1863). *Utilitarianism*. London: Parker, Son and Bourn.

Quinn, W. ([1989a] 1994). 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing', originally in *Philosophical Review*, 98, 1989. Reprinted in Steinbock and Norcross (1994), 355-82.

Quinn, W. (1989b). 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs*, 18, 4: 334-351.

Rowe, W. (1989). 'Causing and Being Responsible for What is Inevitable', *American Philosophical Quarterly*, 26: 153-9.

Sartorio, C. (2004). 'How to be Responsible for Something without Causing It', *Philosophical Perspectives*, 18: 315-36.

Sartorio, C. (2005). 'Causes as Difference-Makers', *Philosophical Studies* 123, 1-2: 71-96.

Sartorio, C. (ms). 'Disjunctive Causes'.

Sinnott-Armstrong, W. (2003). 'Consequentialism', *The Stanford Encyclopedia of Philosophy* (Summer 2003 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2003/entries/consequentialism/>.

Steinbock, B. and A. Norcross (eds.) (1994). *Killing and Letting Die*. New York: Fordham University Press.

Sosa, D. (1993). 'The Consequences of Consequentialism', *Mind*, 102: 101-22.

Thomson, J. ([1976] 1986). 'Killing, Letting Die, and the Trolley Problem', originally in *The Monist*, 59, 1976. Reprinted in William Parent (ed.), *Rights, Restitution, and Risk: Essays in Moral Theory*, Cambridge, Mass: Harvard University Press, 1986, 78-93.

Thomson, J. (2003). 'Causation: Omissions', *Philosophy and Phenomenological Research*, 66: 81-103.

Williams, B. (1973). 'A Critique of Utilitarianism', in J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Yablo, S. (2002). 'De Facto Dependence', *Journal of Philosophy*, 99: 130-48.

Further Reading Suggestions

On consequentialism:

Other classical pieces:

Bentham, J. (1978). *An Introduction to the Principles of Morals and Legislation*. London: T. Payne.

Moore, G. E (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

Anthology of articles on consequentialism:

Scheffler, S. (ed.) (1988). *Consequentialism and Its Critics*. Oxford: Oxford University Press.

On killing and letting die:

Criticisms of the distinction:

Bennett, J. (1995). *The Act Itself*. New York: Oxford University Press.

Rachels, J. ([1975] 1994). 'Active and Passive Euthanasia', *The New England Journal of Medicine*, 292, 1975. Reprinted in Steinbock and Norcross (1994), 112-19.

Defenses of the distinction:

Dinello, D. ([1971] 1994). 'On Killing and Letting Die', *Analysis*, 31, 1971. Reprinted in Steinbock and Norcross (1994), 192-6.

Kamm, F. (1983). 'Killing and Letting Die: Methodological and Substantive Issues', *Pacific Philosophical Quarterly*, 64: 297-312.

On double-effect:

Other classical pieces:

Aquinas, T. ([13th c.] 1988). *Summa Theologica II-II*, Q. 64, art. 7, 'Of Killing', in W. P.

Baumgarth and R. J. Regan (eds.), *On Law, Morality, and Politics*, Indianapolis/Cambridge: Hackett Publishing Co., 1988, 226-7.

Anscombe, E. (1981). 'Mr. Truman's Degree', in *Collected Philosophical Papers of G. E. M. Anscombe*, vol. III. Oxford: Blackwell: 62-71.

Anthology of articles on double-effect:

Woodward, P. A. (ed.) (2001). *The Doctrine of Double Effect: Philosophers Debate a Controversial Moral Principle*. Notre Dame, IN: University of Notre Dame Press.

On moral responsibility:

Accounts of responsibility:

Fischer, J. M. and M. Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Frankfurt, H. (1971). 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy*, 68: 5-20.

Strawson, P. F. (1962). 'Freedom and Resentment', *Proceedings of the British Academy*, 48: 1-25.

On the role of causation in ethics:

Thomson, J. (1990). *The Realm of Rights*. Cambridge, Mass: Harvard University Press, chapter 5. Thomson argues that the notion of consequence presupposed by consequentialism is not causal and offers an alternative account.

McGrath, S. (2003). 'Causation and the Making/Allowing Distinction', *Philosophical Studies*, 114: 81-106. McGrath offers an account of the killing/letting die distinction in terms of causal structures.

Bennett, J. (1995). *The Act Itself*. New York: Oxford, chapter 11. Bennett argues that the causal condition is not an essential component of DDE.

McIntyre, A. (2005). 'Doctrine of Double Effect,' *The Stanford Encyclopedia of Philosophy (Summer 2005 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2005/entries/double-effect/>. *Pace* Bennett, McIntyre regards the causal condition as an essential element of DDE.

Dowe, P. (2001). 'A Counterfactual Theory of Prevention and "Causation" by Omission,' *Australasian Journal of Philosophy*, 79: 216-26. Dowe argues that omissions can only be "quasi"-causes, and that quasi-causation can make us as morally responsible for outcomes as genuine causation does.

* Thanks to Juan Comesaña, Manuel Comesaña, Dan Hausman, and Russ Shafer-Landau for helpful suggestions on an earlier draft.

¹ Note that my focus is “act” consequentialism only, as opposed to, e.g., “rule” consequentialism.

² For an example of the last type, see Feldman (1997). Sosa (1993) argues against the causal view and embraces a view of the second type. Yet another attractive option is to take the consequences of an act to be those aspects of the world that wouldn’t have existed if the act hadn’t been performed (a “counterfactual” view). The causal view only differs from the counterfactual view if causation is not simple counterfactual dependence. But there are persuasive reasons to think it is not [REF 8].

³ For a defense of the opposite view, see Davidson ([1971] 1980). Davidson’s view is that, if the killing is done by shooting, then the act of killing is identical to the act of shooting. Now, the shooting clearly causes the victim’s death; hence the killing causes the victim’s death.

⁴ Someone like David Lewis would not want to say this, but because he thinks that the child’s being an orphan is not an event and thus it cannot enter in causal relations. It is not an event, for Lewis, because it’s “too extrinsic:” it is determined by what happens in a different spatiotemporal region, the one occupied by the child’s parents (Lewis (1986)). But a more liberal conception of events or, alternatively, a more liberal conception of the causal relata would not have this implication.

⁵ I am oversimplifying since the action and the omission could still have different consequences of different types, but we may assume that they don’t, or that they cancel each other out in the

relevant ways. For an objection along similar lines but where consequentialism seems to entail that one should do the act that is intuitively wrong, see Williams (1973).

⁶ I am not objecting to the use of the word ‘cause’ in legal contexts. But, if the word is used in this way in the law, I would say it picks out a different concept from that used in philosophical contexts.

⁷ In fact, I believe that it is in general true that, if c causes e , then the absence of c wouldn’t have caused e . See Sartorio (2005).

⁸ It also depends on what causal consequentialism entails about other cases. A question that arises is: Could causal consequentialism answer the objection illustrated by the Hitler case in the same way? This is less clear. If Hitler wouldn’t have existed if it hadn’t been for the apparently good act, then it is harder to say that this is a case where transitivity fails. But the consequentialist could say that this case is importantly different from the hit man scenario for, in this case, intuitively, the agent *does* contribute (although unknowingly) to Hitler’s atrocities, so it is not too implausible to claim that this fact matters to the morality of the act.

A case that makes more trouble for causal consequentialism is a variant of the hit man case where the hit man threatens to kill a *different* person if I don’t shoot the person that he wants me to kill. If I don’t shoot my victim and the hit man kills his, it seems that I am a cause of that person’s death, for *that* death wouldn’t have occurred otherwise. Still, arguably, if I shoot

my victim, I act impermissibly. Is there a way for the causal consequentialist to address this objection? Although I don't have the space to go into this here, I believe there might be a way. Namely: distinguish between causing a (particular) death and causing deaths (in general), and claim that, although in both scenarios (if I shoot my victim and if I don't) I cause a particular death, I only cause deaths in general if I shoot. I will leave it as an open question whether this suggestion can be developed into a successful answer to the objection.

⁹ Someone might object that, in Trolley too, the one's death is *actually* required for the five to survive (in the sense that, as things stood, the one had to die for the five to survive). But this objection fails. In Trolley, the one's death isn't a cause of the survival of the five. The proposal is not that anything that is actually required is a means, but that anything that is actually a *cause* is a means.

¹⁰ A possibility that I will set aside here is the rejection of causation by omission. Clearly, we can be responsible for things by omission (as in the case of the kidnapper, who is responsible for his victim's death by failing to provide him with food). But, if omissions cannot be causes, then the kidnapper is responsible for his victim's death without having caused it. [REF 19]

¹¹ I focus on Flipper because it is my impression that people's intuitions about Reconnector will be less clear. Unlike Case 1, where the contribution of the two agents is completely symmetrical, here Reconnector's contribution might seem less important than Flipper's: since the train never

gets to run on the segment of main track that he reconnects, one might think that he is off the hook with respect to the death. I will not be concerned with this potential asymmetry here.

¹² In the literature on causation, this type of case is usually called a “switch.” For discussion of switches, see, e.g., Rowe (1989), Yablo (2002), and my (2005).