

## HOW TO BE RESPONSIBLE FOR SOMETHING WITHOUT CAUSING IT\*

Carolina Sartorio  
University of Wisconsin-Madison

### 1. Introduction

What is the relationship between moral responsibility and causation? Plainly, we are not morally responsible for everything that we cause. For we cause a multitude of things, including things that we couldn't possibly foresee we would cause and with respect to which we cannot be assessed morally. Thus, it is clear that causing something does not entail being morally responsible for it. But, does the converse entailment hold? Does moral responsibility require causation? Intuitively, it does: intuitively, we can only be morally responsible for things that we cause.

In this paper I will argue that this intuition is misguided. I will argue that we can be responsible for things that we don't cause, and thus responsibility does not require causation.<sup>1</sup> Moreover, I will argue that this is so *for interesting reasons*. By this I mean two things.

First, I will argue that responsibility does not require causation *even under the assumption that causation by omission is possible*. If, as some philosophers have argued, it were simply impossible to cause something by omission, then, clearly, responsibility would not require causation.<sup>2</sup> For there are things for which we are responsible not in virtue of what we do but in virtue of what we *fail* to do, i.e. in virtue of some omission of ours. So, if causation by omission were impossible, there would be things for which we are responsible without causing them. Following intuition, I will assume that causation by omission is possible. Still, I will argue that being responsible for something does not require causing it.

Second, in order for the thesis that responsibility requires causation to be an interesting target, it must be properly restricted. For, as it stands, the thesis faces a serious problem. The problem arises as follows. The unrestricted version of the thesis says:

**Moral Entails Causal (Unrestricted):** For *any* X, if an agent is responsible for X, it is in virtue of the fact that he caused X, i.e. it is in virtue of the fact that one of his actions or omissions caused X.

The locution “in virtue of the fact that” indicates that, when an agent is responsible for X, he is responsible at least partly *because* he caused X: his causing X partially explains his being responsible for X (“partially” because it cannot be the full explanation, given that causation doesn’t entail responsibility).

Now, the following principle seems true:

- (P) If an agent is responsible for something in virtue of the fact that one of his actions or omissions caused it, then he is responsible for the cause as well.

P is plausible because, if the agent weren’t responsible for the cause, then the fact that the cause brought about the effect wouldn’t have any tendency to explain why the agent is responsible for the effect.

The problem for Moral Entails Causal (Unrestricted) is that, in conjunction with P, it leads to an infinite regress. This emerges as follows. Suppose that we want to hold an agent responsible for an event X. Then Moral Entails Causal (Unrestricted) entails that one of the actions or omissions of the agent caused X. Call that action or omission “Y”. Then P entails that the agent is responsible for Y.<sup>3</sup> But then Moral Entails Causal (Unrestricted) entails that one of the actions or omissions of the agent caused Y. Call that action or omission “Z”. And so on. This is a problem because it means that, in order for an agent to be responsible for something, he must be responsible for an *infinite* number of things. And, in principle, it is not easy to see how this could be so.

However, the thesis that responsibility requires causation may be easily restricted in a way that avoids this problem. In particular, we may restrict it to outcomes in the external world (events or states of affairs, such as a person’s death or a person’s being harmed). The thesis that responsibility for *outcomes* requires causation is widespread among philosophers.<sup>4</sup> And it is no mystery why this is so. Clearly, we can only be responsible for what happens in the external world if we are hooked up to the world in some way. Now, the only way in which it seems that we could be hooked up to the world is by means of our actions and omissions. And the only way in which our actions and omissions could hook us up to the world seems to be by means of what they cause. Thus, the natural thought is that, if we are responsible for an outcome, it must be because our actions or omissions caused the outcome.

Moreover, we often seem to put this intuitive idea to work in the following way. Imagine that we believe that a person is responsible for a certain outcome. However, we then find out that nothing that the person did or failed to do

caused the outcome. Then it is likely that we will abandon our belief that the person is responsible for the outcome. Imagine, for instance, that a sniper and I willingly fire our guns at the same time in my enemy's direction. My enemy dies and the autopsy reveals that only the sniper's bullet reached him and killed him. Then we will conclude that I am not responsible for my enemy's death, although I am responsible for trying to kill him. The reason why I am not responsible for the death is, intuitively, that I didn't cause it, even if I tried. In other words, the reason why I am not responsible for the death seems to be that my firing my gun, the only thing I did that could have made me responsible for the death, wasn't a cause of it.

In short, under the assumption that there is causation by omission, the following principle seems very plausible:

**Moral Entails Causal:** If an agent is responsible for an outcome, it is in virtue of the fact that he caused it (some action or omission of his caused it).

In the first part of the paper I will argue that Moral Entails Causal is false. Then, in the second part of the paper, I will try to do some rebuilding. I will address the questions that the first part naturally gives rise to, namely: if we can be responsible for outcomes without causing them, then, does this mean that there is no connection between responsibility and causation? How can we be responsible for what goes on in the world without being causally connected to the world by means of our actions and omissions? I will make an alternative proposal about the relation between responsibility for outcomes and causation, and I will argue that the alternative proposal is, on reflection, as plausible and as helpful as Moral Entails Causal seemed to be. Finally, I will draw some implications of this view for the voting problem: the problem of accounting for the rationality of casting a vote in an election, even if, in general, a single vote makes no difference to the outcome of the election.

## **2. The argument against the received view**

Imagine the following situation. There was an accidental leak of a dangerous chemical at a high-risk chemical plant, which is on the verge of causing an explosion. The explosion will occur unless the room containing the chemical is immediately sealed. Suppose that sealing the room requires that two buttons—call them “A” and “B”—be depressed at the same time  $t$  (say, two seconds from now). You and I work at the plant, in different rooms, and we are in charge of accident prevention. Button A is in my room, and button B is in yours. We don't have time to get in touch with each other to find out what the other is going to do; however, we are both aware of what we are supposed to do. As it turns out, each of us independently decides to keep reading his magazine instead of depressing his button. The explosion ensues.

Now consider the following variant of the case. Again, button A is in my room, and I fail to depress it. This time, however, there is no one in the room containing button B; instead, a safety mechanism has been automatically set to depress B at *t*. When the time comes, however, B becomes stuck while being up. Just as in the original case, then, neither button is depressed and the explosion occurs. Call the two cases “Two Buttons” and “Two Buttons-One Stuck,” respectively. The cases differ in the respect that, in Two Buttons, B isn’t depressed because you decided not to depress it, whereas, in Two Buttons-One Stuck, it isn’t depressed because it got stuck.

I will argue that Two Buttons is a case of responsibility without causation, and thus it is a counterexample to Moral Entails Causal. My argument will take the following form:

- (1) I am responsible for the explosion in Two Buttons.
- (2) My failure to depress A didn’t cause the explosion in Two Buttons-One Stuck.
- (3) If my failure to depress A didn’t cause the explosion in Two Buttons-One Stuck, then it didn’t cause it in Two Buttons.
- (4) Therefore, my failure to depress A didn’t cause the explosion in Two Buttons. (From (2) and (3))
- (5) No other action or omission of mine caused the explosion in Two Buttons.
- (6) Therefore, Moral Entails Causal is false. (From (1), (4) and (5))

In other words, I will argue that I am responsible for the explosion in Two Buttons but nothing I did or failed to do caused it. In particular, my failure to depress A didn’t cause it. And I will argue that my failure to depress A didn’t cause the explosion in Two Buttons by arguing that it didn’t cause it in Two Buttons-One Stuck and that my causal powers with respect to the explosion are the same in the two cases.

I take (5) to be clearly true;<sup>5</sup> thus, I won’t argue for it here. In the following sections, I take up premises (1) through (3) in turn. First, however, a note on the methodology is in order. I will be arguing that we should accept the premises in my argument because the price of rejecting them is very high. Now, it might well be that the price of giving up Moral Entails Causal, the thesis that my argument attacks, is *also* high (especially since, as I have granted, Moral Entails Causal is an intuitively plausible and fruitful principle). If so, we should do whatever comes at the smallest price. And, how are we going to know what this is? Here is where the positive proposal of this paper steps in. I will argue that the price of giving up Moral Entails Causal is actually not high at all, for an alternative principle about the relation between causation and responsibility is at least as plausible and at least as fruitful. As a result, we shouldn’t have qualms about abandoning Moral Entails Causal.

### 3. Argument for the first premise

In this section I will argue for premise (1):

(1) I am responsible for the explosion in Two Buttons.

I find (1) intuitively true.<sup>6</sup> In addition, there is a persuasive argument in support of this intuition. Briefly, the argument goes as follows. If we were to reject (1), then we would have to say that Two Buttons is a case of moral luck. But it would be wrong to count situations like Two Buttons as situations of moral luck. Hence, we should accept (1).

Let me explain. A case of *moral (good) luck* is a case where an agent, who behaves in a way that is generally conducive to a certain type of harm, is relieved of any responsibility for the harm (even if the harm ensues) thanks to the obtaining of some circumstances that are outside of the agent's control.<sup>7</sup> For instance, if I fire my gun at my enemy and the bullet is deflected by a gust of wind, but at the same time he is struck by a lightning and dies, then I am not responsible for my enemy's death, even if I acted badly (in a way that could very easily have caused the person's death). Thus, this is a case of moral luck. Now, if I weren't responsible for the explosion in Two Buttons, then Two Buttons would be a case of moral luck. For it would be a case where I am not responsible for the ensuing harm even if I acted in a morally unacceptable way that is generally conducive to that type of harm. In addition, I would not be responsible for the harm due to the obtaining of some circumstances that were out of my control, namely, the fact that B also wasn't depressed.

However, I don't think that we are prepared to accept this as a genuine kind of moral luck. The first thing to notice is that, if we were to say that I am not responsible for the explosion in Two Buttons, then we would have to say the same about you: we would have to say that I am not responsible because B wasn't depressed, and you are not responsible because A wasn't depressed. Now, B wasn't depressed because *you* failed to depress it, and A wasn't depressed because *I* failed to depress it. Thus, if we said that Two Buttons is a case of moral luck, we would have to say that, for each of us, the fact that the other *also* behaved in a morally unacceptable way (that is generally conducive to a certain type of harm) is enough to relieve him of responsibility for the harm. But, in Two Buttons, the harm occurred in virtue of the fact that we behaved badly: had both of us done the right thing, the harm wouldn't have occurred. Thus, claiming that Two Buttons is a case of moral luck amounts to claiming that two wrongs that are generally conducive to a certain type of harm can neutralize each other in circumstances where they are jointly responsible for the occurrence of the harm. And this seems wrong.

In other words, we regard the situation in Two Buttons as one where a purely *human* failure took place, and thus we want to assign blame for what

happened to the moral agents involved. The fact that the human failure is traceable to more than one human being does not mean that the agents are relieved of responsibility; rather, it means that they *share* responsibility for the bad outcome, just as the members of a gang share responsibility for a robbery.<sup>8</sup>

Contrast this with Two Buttons-One Stuck. In Two Buttons-One Stuck, B wasn't depressed due to a purely mechanical—not human—failure. Intuitively, mechanical failures are the kind of thing that can give rise to moral luck. Intuitively, in Two Buttons-One Stuck, even though I thought that I could have prevented the explosion by depressing A, and even if I acted badly in failing to depress A, I was lucky. I was lucky because B got stuck at the time at which I should have depressed A. The fact that B was stuck seems to exempt me from responsibility for the explosion. In other words, Two Buttons-One Stuck strikes us as a typical case of moral luck, where some natural phenomenon that is outside my control takes away my responsibility for the outcome.<sup>9</sup>

Two Buttons-One Stuck is similar to cases that philosophers have discussed in the context of the debate over whether one can be responsible for outcomes that one couldn't have prevented. Here is an example: I am walking by the beach when I see that a child is drowning. I think I could prevent his death, but I deliberately refrain from jumping in to attempt the rescue. The child drowns. Unbeknownst to me, however, there was a patrol of hungry sharks in the water that would have attacked me as soon as I jumped in, and thus I couldn't have saved the child. Am I responsible for the death of the child under those circumstances? It seems not: it seems that I am responsible for not trying to save the child, but not for his death.<sup>10</sup> Similarly, it seems that, in Two Buttons-One Stuck, I am responsible for not trying to prevent the explosion, but not for the explosion itself.

In sum, there seems to be an interesting moral difference between Two Buttons and Two Buttons-One Stuck: I am responsible for the explosion in Two-Buttons, but not in Two Buttons-One Stuck.<sup>11</sup> Now, what explains this difference? In both cases, I couldn't have prevented the explosion by depressing A, since the other button wasn't going to be depressed (in one case, because it got stuck; in the other case, because you failed to depress it). Why, then, am I responsible for the explosion in one case but not in the other? I will return to this question in section 7. As we will see in the next two sections, it is *not* because I cause the explosion in one case and not in the other, since, as I will be arguing, there is no such causal difference between Two Buttons and Two Buttons-One Stuck.

#### 4. Argument for the second premise

In this section I will argue for premise (2):

- (2) My failure to depress A didn't cause the explosion in Two Buttons-One Stuck.

I will argue that we should endorse (2), or else we would be committed to much more causation by omission than we are prepared to accept.

Imagine that we said that my failure to depress A did cause the explosion in Two Buttons-One Stuck. To say that it did is to say that it caused it even if B was stuck and, hence, even if depressing A wouldn't have prevented the explosion. Had I depressed A *and had B not been stuck*, then the explosion would have been prevented, but my depressing A wouldn't have been sufficient to prevent it. Still, it would have been a cause. Now, if this is so, we should probably say that an agent's failure to act in the relevant way caused the outcome in all of the following cases. If a child is drowning but there are sharks in the water that would have thwarted a rescue attempt, we would have to say that I caused the death of the child by failing to jump into the water to rescue him, when I couldn't have saved him, given the presence of the sharks. After all, had I jumped into the water to save him *and had there been no sharks in the water*, I would have saved him. Similarly, we would have to say that a doctor's failing to operate on a patient with a tumor caused the patient's death, when he couldn't have saved him, for the tumor was too deep into the patient's brain and thus couldn't be removed. After all, had the doctor operated *and had the tumor not been so deep into the brain*, the patient would have lived. Notice, in particular, that we would have to say this even if the doctor was fully aware of the fact that the tumor couldn't be removed (and, in the drowning case, even if I was fully aware of the presence of the sharks), for an agent's epistemic state is irrelevant to the causal powers of his actions and omissions.

More generally, take any outcome that I couldn't have prevented by acting in a certain way, given the existence of an obstacle or impediment. Still, had I acted that way *and had the obstacle been absent*, I would have prevented the outcome. Hence, if we were to say that I am a cause of the explosion in Two Buttons-One Stuck, then we would have to say that I caused the outcome in each of those cases. Moreover, whether there is one or more than one obstacle couldn't possibly make a difference to my causal powers. Hence, if we were to say that I am a cause of the explosion in Two Buttons-One Stuck, we would have to say that I am a cause of any outcome when there were many obstacles; after all, had those obstacles been absent and had I acted in the relevant way, I would have prevented the outcome. (Imagine that, in addition to the sharks, there are other obstacles to my saving the drowning child: a big concrete wall in the water, strong currents, explosives, and so on; it is very implausible to believe that my failure to jump in was a cause of the child's death when there were all those obstacles to my saving him.)

Presumably, the problem is serious enough so that, if we were to pursue this route, we would be committed to saying that *any* omission is a cause of *any* (contingent) outcome that follows. For take an arbitrary outcome, say, the death of my friend's plant in Argentina today, and an apparently completely unrelated omission, say, my failure to carry an umbrella to work yesterday in Madison, Wisconsin. Now consider this other omission: Larry's failure to have

the conditional disposition to get on a plane to Argentina and water my friend's plant before it died if I had carried an umbrella to work yesterday. Had I carried an umbrella to work yesterday *and had Larry had such a disposition*, then the plant wouldn't have died today. Hence, if we were to say that I am a cause of the explosion in Two Buttons-One Stuck, we would probably have to say that my failure to carry an umbrella to work yesterday was a cause of the plant's death, and, in general, that any omission of mine is a cause of anything that followed.

Notice that the problem is much more serious than a commonly noted problem concerning causation by omission. It has been noted that accepting the possibility of causation by omission leads to the existence of too much causation. For instance, if we were to say that my failure to water a plant that I promised to water is a cause of its death, then we would probably also have to say that the *Queen of England's* failure to water the plant is a cause of its death (because it is also true of the Queen of England that, had she watered the plant, the plant would have survived). This is a problem because we wouldn't ordinarily count the Queen of England's failure to water the plant as a cause of the plant's death. Now, the problem generated by counting my failure to push the button as a cause in Two Buttons-One Stuck is much more serious than the "Queen of England problem." The Queen of England problem would have us say that every failure to water the plant (including the Queen of England's failure) is a cause of the plant's death. By contrast, the problem I have been discussing would have us say that *any* omission whatsoever is a cause of any ensuing outcome, and this includes many omissions that are, on the face of it, intuitively unrelated to the occurrence of the outcome (unlike a failure to water a plant and the plant's death, which are importantly related to each other).

Another important difference between the problem that occupies us and the Queen of England problem is the following. The Queen of England problem would have us say that many failures, including some unexpected failures, are causes. But these would not be more than *joint* causes. This is to say, each one of those failures would contribute only part (a very small part) of what was required for the outcome to happen: in order for the plant to die, I had to fail to water it, the Queen of England had to fail to water it, and everyone else had to fail to water it. By contrast, if button A's not being depressed were a cause of the explosion in Two Buttons-One Stuck, then, by the same token, B's not being depressed would also be a cause. However, they would be more than joint causes. Recall that one's button not being depressed was *sufficient* for the explosion to occur. So, if the buttons not being depressed were causes, they would not be joint causes but "overdeterminers": each, independently of the other, would have contributed (not part of but) all of what was required for the explosion to occur. As a result, if we were to say that my failure to depress A is a cause of the explosion in Two Buttons-One Stuck, then we would probably have to say, not only that any omission *causes* any outcome that ensues, but also that every omission *overdetermines* every outcome that ensues. And this is very implausible.<sup>12</sup>



I conclude that rejecting premise (2) would have many extremely implausible and undesired consequences. Therefore, we should accept it.

## 5. Argument for the third premise

Finally, let us turn to the third premise in my argument:

- (3) If my failure to depress A didn't cause the explosion in Two Buttons-One Stuck, then it didn't cause it in Two Buttons.

My argument for (3) will be based on the fact that the two cases, Two Buttons and Two Buttons-One Stuck, are relevantly similar. To be sure, there are differences between them; as I have briefly indicated, those differences are enough to ground a moral difference between the cases. But I will argue that the differences that there are could not plausibly be viewed as mattering *causally*.

The cases have been laid out in such a way that the only important difference between them is that a *person* is in control of button B in one case but a *mechanism* is in control in the other. In both cases, B isn't depressed at the required time, but, in Two Buttons, it is because you failed to depress it, whereas, in Two Buttons-One Stuck, it is due to a mechanical failure. Hence rejecting (3) would amount to claiming that whether there is a causal connection between my failure to depress A and the explosion can depend on whether *a person*, as opposed to an unconscious mechanism of some sort, is in the other room. But this is highly implausible. That is, it is highly implausible to believe that the mere fact that a person is in the other room, as opposed to a machine that behaves in relevantly similar respects as the person does, might make a difference to my causal powers. The fact that B wasn't depressed certainly matters to my causal powers, given that B's being depressed was necessary to prevent the explosion. But it seems that whether B wasn't depressed as a result of a person's failing to depress it or as a result of a mechanical failure of some sort simply shouldn't be relevant to whether I caused the explosion by failing to depress A.

Let me illustrate with an example. In order for the example to be sufficiently analogous, I suggest that we look at a purported case of overdetermination. (The reason for choosing this type of example is that, as we saw in the last section, cases like Two Buttons-One Stuck have the basic structure of an overdetermination case in the sense that, were we to say that there is causation, we would thereby have to say that there is overdetermination, for each of the individual omissions is independently sufficient for the outcome). Take a case of two rocks simultaneously hitting a window and making it shatter, and imagine that I threw one of the rocks. Now, suppose that we are trying to establish whether my throwing my rock was a cause of the shattering. Would it matter, for these purposes, whether the other rock was thrown by another person or by an unconscious mechanism (say, an automated catapult)? Clearly

not. Whether a person or a catapult threw the other rock seems completely irrelevant to the causal powers of my throw. What does seem to matter is *whether* the other rock impacted the window (or whether it was just my rock that impacted it), and *how* it impacted it (in particular, whether it made any important difference to the shattering of the window, or whether my rock was responsible for the major crack that ended in its shattering). But whether all this happened because a sentient being or an unconscious mechanism threw the other rock is simply irrelevant to whether my throw caused the shattering.

Similarly, it seems that whether a person was in charge of B in the other room, or whether a mechanism was, should simply be irrelevant to whether my failure to depress A caused the explosion. What does seem relevant is whether B was depressed (if it had been depressed, then my failure to depress A would have been a cause), and what effect B's not being depressed had on the explosion (if depressing only A had been sufficient to prevent the explosion, then, again, my failure to depress A would have been a cause). But whether B wasn't depressed as a result of a human failure or as a result of a mechanical failure seems irrelevant to whether my failure to depress A was a cause of the explosion.<sup>13</sup>

I have argued that the differences between Two Buttons and Two Buttons-One Stuck do not matter causally. I will conclude my defense of (3) by drawing attention to the fact that the main existing theories of causation are likely to regard the two cases as causally on a par: they are likely to entail that, either my failure to depress A caused the explosion in both cases, or in neither case (although different theories might disagree about whether it is a cause in both cases or in neither case). To see this, let us quickly review the two main categories of theories of causation.

Traditionally, theories of causation are classified into "regularity" theories and "counterfactual" theories. Very roughly, and in its simplest version, a regularity theory deems something a cause when it is sufficient, in the circumstances, and given the laws, for the occurrence of the effect. And, also very roughly, and in its simplest version, a counterfactual theory deems something a cause when the effect counterfactually depends on it, i.e. had the cause not occurred, the effect wouldn't have occurred. A *regularity* theory is likely to entail that my failure to depress A is a cause of the explosion in *both* cases, Two Buttons and Two Buttons-One Stuck. For my failure to depress A was sufficient, in the circumstances, and given the laws, for the explosion. Whether there is a person or a mechanism in the other room is simply irrelevant to the fact that my failure to depress A was sufficient, in the circumstances and given the laws, for the explosion. A *counterfactual* theory, by contrast, is likely to entail that my failure to depress A is a cause in *neither* case. For, given that the explosion would only have been prevented by depressing both buttons, and given that the other button wasn't depressed, the explosion would still have occurred if I had depressed A. So the explosion doesn't counterfactually depend on my failure to depress A, and thus a counterfactual theory would not count my failure to depress A as a cause of the explosion. Again, whether the other

button wasn't depressed due to a human or a mechanical failure is simply irrelevant to the fact that the explosion doesn't counterfactually depend on my failure to depress A.

Despite their differences, then, both regularity theories and counterfactual theories are likely to entail that the two cases are causally on a par. Naturally, there are many varieties of both regularity and counterfactual theories of causation, and I do not intend for this brief sketch of theories of causation to span them all. However, it does serve as an indication that the kinds of factors that are generally considered to be causally relevant are not the kinds of factors that distinguish Two Buttons from Two Buttons-One Stuck. As a result, my claim that the two cases are causally on a par does not seem to be particularly controversial.

This concludes my discussion of the premises of my argument against Moral Entails Causal. To sum up, my argument has been the following:

- (1) I am responsible for the explosion in Two Buttons.
- (2) My failure to depress A didn't cause the explosion in Two Buttons-One Stuck.
- (3) If my failure to depress A didn't cause the explosion in Two Buttons-One Stuck, then it didn't cause it in Two Buttons.
- (4) Therefore, my failure to depress A didn't cause the explosion in Two Buttons. (From (2) and (3))
- (5) No other action or omission of mine caused the explosion in Two Buttons.
- (6) Therefore, Moral Entails Causal is false. (From (1), (4) and (5))

The following emerges from the discussion so far. Let us coin the phrase "causal luck" to refer to the following phenomenon: two factors that would have been causally efficacious if they had acted alone cancel each other's causal powers out when they occur simultaneously. Then what Two Buttons and Two Buttons-One Stuck show is that *causal* luck is more common than *moral* luck. At least in the case of omissions, causal luck obtains when, given that the two factors occur simultaneously, a certain abstract dependence relation that would otherwise have existed between each of the factors and the effect is absent. Moral luck, by contrast, is sensitive to other features of the situation, in particular, it is sensitive to whether what breaks the dependence between each of the factors and the effect is another *moral agent*. As a result, causal luck can occur without moral luck. Hence, there can be responsibility without causation.<sup>14</sup>

## 6. Towards the new view

What is the relationship between responsibility and causation, if it is not entailment? Before addressing this question, I will consider a different question

that arises naturally in light of the preceding discussion. As we will see, this will also serve the further purpose of helping us rethink the relationship between responsibility and causation.

The question that arises in light of the preceding discussion is the following. If, as I have argued, my failure to depress A did not cause the explosion in Two Buttons (or in Two Buttons-One Stuck, but let us focus on Two Buttons) and, by similar reasoning, your failure to depress B didn't cause it either, then *what* did? Naturally, the chemicals having leaked out of the place where they were stored did, but this answer isn't fully satisfying: the buttons (in virtue of their not being depressed) seemed to have had something to do with the explosion too. The explosion occurred (partly) *because* the buttons weren't depressed (because the preventive mechanism constituted by the pair of buttons wasn't activated). Hence, the reconstruction of the causal history of the explosion would seem incomplete if it didn't make any reference to the buttons whatsoever.<sup>15</sup> The question is, then, how should this gap in the causal history of the explosion be filled?

I will suggest that some *other* condition, not my failure to depress A, and not your failure to depress B, although one that is closely related to both, caused the explosion in Two Buttons. What is this other condition? To see what it is, consider first an example involving just one agent. Imagine that an orchestra has delivered a wonderful performance. At the end of the concert, I am expected to clap. Instead, I remain completely still. As a result, Jim forms the belief that I was rude. What caused Jim's belief that I was rude? Clearly, it was my failure to clap. What is my failure to clap? It is my failure to simultaneously move my left hand and my right hand in particular ways. My failure to simultaneously move my left hand and my right hand in particular ways obtains just in case *either* I fail to move my left hand in particular ways at the required time *or* I fail to move my right hand in particular ways at the required time, or both. Had I moved just my left hand, I wouldn't have clapped, and thus Jim would still have thought that I was rude. Had I moved just my right hand, I wouldn't have clapped either, and thus Jim would have thought that I was rude too. Jim would only have failed to think that I was rude if I had moved *both* of my hands in the ways that clapping requires.<sup>16</sup>

Let us represent the different conditions schematically. Let  $F(L)$  be my failure to move my left hand (in the required way, at the required time) and let  $F(R)$  be my failure to move my right hand (in the required way, at the required time). Then we may represent my failure to clap as  $F(L \wedge R)$ .  $F(L \wedge R)$  is my failure to *both* move my left hand and my right hand (in the required way, at the required time), and it obtains whenever at least one of the individual failures obtains (i.e., it is equivalent to the disjunction of the individual failures,  $F(L) \vee F(R)$ ).  $F(L \wedge R)$  should be distinguished from each of the individual failures,  $F(L)$  and  $F(R)$ , as well as from the condition that results from conjoining the two,  $F(L) \wedge F(R)$ , which obtains just in case *both* individual failures obtain.  $F(L) \wedge F(R)$  entails both of  $F(L)$  and  $F(R)$ , since every world where  $F(L) \wedge F(R)$

obtains is a world where  $F(L)$  obtains and also a world where  $F(R)$  obtains. In turn, each of  $F(L)$  and  $F(R)$  entails  $F(L \wedge R)$ , since every world where  $F(L)$  obtains is a world where  $F(L \wedge R)$  obtains and every world where  $F(R)$  obtains is a world where  $F(L \wedge R)$  obtains.

My claim, then, is that  $F(L \wedge R)$  caused Jim's belief that I was rude.  $F(L \wedge R)$  obtains in every world where I fail to move *at least one* hand. In all and only those worlds, Jim would have believed that I was rude. This is a *prima facie* reason to believe that  $F(L \wedge R)$  is a cause of Jim's belief.<sup>17</sup>

I submit that the situation in Two Buttons is analogous. Two Buttons is a case with essentially the same structure as that of the clapping case, with the only difference that it involves two agents instead of one. In Two Buttons, the explosion would only have been prevented if we had simultaneously depressed A and B at  $t$ . As a matter of fact, we didn't simultaneously depress A and B at  $t$ . I submit that *our failure to simultaneously depress A and B at  $t$*  caused the explosion. What is our failure to simultaneously depress A and B at  $t$ ? It is the condition that obtains just in case *either* I fail to depress A at  $t$ , *or* you fail to depress B at  $t$ , or both. This condition obtains in the actual world given that both of us failed to depress our buttons, but it also obtains in worlds where only one of us fails to depress his button.

If  $F(A)$  is my failure to depress A and  $F(B)$  is your failure to depress B, our failure to simultaneously depress A and B at  $t$  can be represented as  $F(A \wedge B)$ . Just as in the clapping case,  $F(A \wedge B)$  should be distinguished from each of our individual failures,  $F(A)$  and  $F(B)$ , as well as from  $F(A) \wedge F(B)$ , the condition that obtains just in case *both* of us fail to depress our buttons. Instead,  $F(A \wedge B)$  obtains whenever *at least one* of us fails to depress his button (i.e., it is equivalent to  $F(A) \vee F(B)$ ). Also, just as in the clapping case,  $F(A \wedge B)$  is entailed by  $F(A)$  and  $F(B)$ . And, finally, just as in the clapping case, there is *prima facie* reason to believe that it is a cause of the outcome, the explosion. Why? For the same reason that my failure to clap is likely to be a cause of Jim's forming the belief that I was rude in the clapping case, namely, the fact that, given the circumstances, the explosion occurs in all and only the worlds in which  $F(A \wedge B)$  obtains. These include worlds where both of us fail to depress our buttons, but also worlds where only one of us does.

Let me sum up. The question that I wanted to address in this section was this: If, as my argument against Moral Entails Causal suggests, my failure to depress A did not cause the explosion in Two Buttons, and the same goes for your failure to depress B, then what did? Certainly, the two buttons had something to do with the explosion's coming about. My reply is that our failure to simultaneously depress A and B did. This is a condition that obtains whenever at least one of us fails to depress his button.

The question will surely arise: Is it really possible to hold, as I am suggesting, that our failure to simultaneously depress A and B caused the explosion, but neither of our individual failures, which entail it, did? More generally, is the following scenario really possible: X entails Y, Y causes E, but X *doesn't* cause

E? I think that there is good reason to believe it is possible. Consider the following case. Suzy just learned that people are mortal. In particular, she just learned that Grandpa, who she adores, is going to die someday, and this made her cry. Now, imagine that, unbeknownst to Suzy, Grandpa just died of a heart attack. The fact that he died entails the fact that he was mortal. But the fact that Grandpa died didn't cause Suzy to cry: she didn't cry because Grandpa died (she doesn't even know that he died), but because he is mortal. So it is possible for X to entail Y, for Y to cause E, but for X *not* to cause E.

Now, how is this going to help us figure out the relationship between responsibility and causation? I turn to this question in the following section.

## 7. Causation as the vehicle of transmission of responsibility

How can I be responsible for the explosion in Two Buttons, if my failure to depress A didn't cause it? I suggest the following. In Two Buttons, I am responsible for the explosion, not in virtue of the fact that my failure to depress A caused it (since it didn't), but in virtue of the fact that *something for which I am responsible* caused it. In other words, I am responsible for the explosion because there is a cause of the explosion for which I am responsible.<sup>18</sup> What is this cause of the explosion for which I am responsible? I submit it is our failure to simultaneously depress A and B. In the last section, I argued that this condition is a cause of the explosion in Two Buttons. Now I will argue that I am responsible for it.<sup>19</sup> It will then follow that something for which I am responsible caused the explosion in Two Buttons.

The reason why I think we should say that I am responsible for our failure to simultaneously depress A and B in Two Buttons is similar to the reason why we should say that I am responsible for the explosion. As we have seen in section 3, we should say that I am responsible for the explosion in Two Buttons (and so are you) or else there would be things for which no one is responsible but that depend exclusively on the morally unacceptable behavior of some moral agents. And this is implausible. Now, our failure to simultaneously depress A and B depends exclusively on the morally unacceptable behavior of some moral agents, namely, you and me. Hence we should say that I am responsible for our failure to simultaneously depress A and B, and so are you. In other words, we should say that each of us is responsible for this failure. If so, I am responsible for a cause of the explosion in Two Buttons.

By contrast, I am presumably responsible for none of the causes of the explosion in Two Buttons-One Stuck. First, as we have seen, my failure to depress A didn't cause the explosion. Hence, even if I am responsible for my failure to depress A, such failure wasn't a cause of the explosion. And, second, just as I am not responsible for the explosion given that B was stuck (as I have pointed out in section 3), I am presumably also not responsible for the failure of the two buttons to be simultaneously depressed. Figuring out precisely *why* this is so would require an in-depth investigation of the intriguing phenomenon of

moral luck, a task that I cannot pursue here. But the fact remains that B wasn't depressed due to a mechanical failure, not a human failure, and somehow this seems to take away my responsibility for the fact that the two buttons weren't simultaneously depressed. Hence, even if the failure of the two buttons to be simultaneously depressed caused the explosion (as we saw in the last section), I am not responsible for that failure. It seems, then, that I am not responsible for any of the causes of the explosion in Two Buttons-One Stuck.

In sum, I suggest the following principle about the relationship between causation and responsibility (a principle that helps explain the moral difference between Two Buttons and Two Buttons-One Stuck):

**Causal Transmits Moral:** If an agent is responsible for an outcome, then it is in virtue of the fact that the agent is responsible for something that caused the outcome.<sup>20</sup>

In other words, according to Causal Transmits Moral, if I am responsible for an outcome, then it doesn't have to be the case that something that I did or failed to do caused the outcome, but it does have to be the case that I am responsible for one of the outcome's causes. The outcome's cause that I am responsible for might be an action or omission of mine, but it can also be, as in Two Buttons, the collective behavior of a group of agents.

Note that Causal Transmits Moral is restricted to *outcomes*, just as Moral Entails Causal was. This prevents Causal Transmits Moral from leading to an infinite regress, since, if it weren't thus restricted, it would follow that, in order for an agent to be responsible for something, he would have to be responsible for one of its causes, and for one of the cause's causes, and so on. Given that it is restricted to outcomes, Causal Transmits Moral does *not* entail that one must be responsible for an infinite number of causes of an outcome in order to be responsible for the outcome.<sup>21</sup>

Independently of the two cases that have occupied us, Two Buttons and Two Buttons-One Stuck, Causal Transmits Moral is an intuitively plausible principle about the relation between responsibility (for outcomes) and causation. On the face of it, we can only be responsible for an outcome if we are responsible for one of its causes. For instance, it seems that I wouldn't be responsible for a death by shooting unless I were responsible for the bullet's piercing the person's heart, or for some other contributing cause, such as the fact that the person was standing in front of the gun.

Moreover, not only does Causal Transmits Moral seem plausible, but it also appears to be as fruitful as Moral Entails Causal seemed to be. The paper started out with the remark that Moral Entails Causal appears to explain my lack of responsibility in cases like the following: I shoot at my enemy and miss; however, at the same time, a sniper shoots the bullet that kills him. Intuitively, I am not responsible for my enemy's death, and Moral Entails Causal seemed to



explain why: I could only be responsible for his death in virtue of having shot a bullet at him, but my bullet didn't cause his death; hence, I am not responsible for his death. I have argued that Moral Entails Causal is false. However, its substitute Causal Transmits Moral could explain equally well why I am not responsible in this case. Given that I am not responsible for any of the causes of the death (e.g., the sniper's shooting, or my enemy's standing within the sniper's shooting range), it follows from Causal Transmits Moral that I am not responsible for the death.

We have seen that Causal Transmits Moral is an initially plausible way of understanding the relation between responsibility and causation. In addition, it is an improvement over Moral Entails Causal in that it does not overlook cases like Two Buttons, where, as I have argued, an agent is responsible for an outcome without causing it. Finally, it is as useful as Moral Entails Causal seemed to be in that it successfully accounts for the lack of responsibility of an agent for an outcome in cases where he is not responsible for any of its causes. I conclude that there is good reason to believe that Causal Transmits Moral successfully captures the relation between responsibility and causation.

## 8. Conclusions and implications for the voting problem

What lessons should we draw from all this? One lesson we should draw is the following. Agents are responsible for what happens in the external world in virtue of how they interact with it—by means of their actions and omissions. *However*, their actions and omissions can make them responsible for things by virtue of more than their causal powers: in particular, they can make agents responsible for things by virtue of the causal powers of larger collective behaviors of which those actions and omissions are parts. That is the type of scenario depicted by Two Buttons.<sup>22</sup>

Another lesson we should draw concerns the general way in which to regard the concepts of responsibility and causation in connection with each other. There is a strong temptation to regard causation as a *condition* on responsibility. Quite generally, being responsible for an outcome tends to be associated with, roughly, intentionally (or negligently) causing the outcome.<sup>23</sup> If, as I have claimed, Moral Entails Causal is false, then this is a mistake: one can be responsible for an outcome without even causing the outcome. It follows that we shouldn't view the relation between responsibility and causation as a kind of entailment relation. How should we view it, then? If, as I have argued, Causal Transmits Moral is true, then causation should be rather viewed as the *vehicle of transmission* of responsibility. This is to say, in order for an agent to be responsible for an outcome, there must be a causal link between an earlier thing (event, state of affairs, action, omission, etc.) and the outcome along which the responsibility of the agent was transmitted.<sup>24</sup>

I will conclude by drawing attention to some consequences of this view for the voting problem. As I will understand it, the voting problem is the problem of



explaining why citizens should vote even if, in the vast majority of cases, each particular vote does not make a difference to the outcome of the election. From a rational-choice perspective, it would seem that it is generally not rational for citizens to vote, given that the expected benefit of voting does not outweigh the expected cost. Alvin Goldman has argued for a *causal responsibility* solution to this problem, according to which, even if it were true that there are no *prudential* reasons to vote, there would still be *moral* reasons (Goldman (1999)). According to Goldman, the moral reasons to vote arise from the fact that a citizen can be a *cause* of the outcome of an election (and thus, can deserve credit or blame for it) by voting or abstaining, even if whether he votes or abstains makes no difference to the outcome.

Suppose that there are two candidates, Good and Bad, where Good is much more preferable than Bad. Suppose that Bad wins by some margin, and that several people abstain (imagine that, had all the abstainers voted for Good, Good would have won). According to the causal responsibility approach, everyone that voted for Bad was a cause of Bad's victory. In addition, everyone that abstained was a cause of Bad's victory (although, on Goldman's view, the causal responsibility of an abstainer is somewhat less than the causal responsibility of someone that voted for Bad). In particular, take someone, Lazy, who would have voted for Good but didn't vote at all. On the causal responsibility approach, Lazy is a cause of Bad's victory. By contrast, if Lazy had voted for Good, he wouldn't have been a cause of Bad's victory. Since one ought to avoid causing harm, this account gives potential voters a moral reason to vote for the better candidate.

Goldman supports his claim that someone like Lazy causes Bad's victory by offering a "vectorial" model of causation (a model that applies to what he calls "vectorial causal systems," of which electoral systems are prime examples), on which this is true. But, if my arguments in this paper are sound, then this model of causation is not right. My diagnosis of Two Buttons suggests that Lazy does *not* cause the outcome of the election by abstaining. (For, if I don't cause the explosion in Two Buttons by failing to depress my button, it seems that Lazy doesn't cause Bad's winning the election by abstaining either.)<sup>25</sup> Hence, if my arguments in this paper are sound, it follows that the causal responsibility approach does not solve the voting problem: we still need an account of why citizens should vote.

At the same time, however, the view that I have defended in this paper suggests an alternative way in which to try to solve the voting problem. I have argued that, in Two Buttons, although my individual omission doesn't cause the explosion, our simultaneous failure to depress both buttons does. Election cases might be similar in that, although certain individual omissions don't cause the good candidate to lose, a more complex condition does. Election cases are not completely analogous to Two Buttons because, in a typical election case, there is no *single* combination of individual votes that would have resulted in a certain candidate's winning. However, the cause in an election case could be a more

abstract condition that is realizable by different combinations of individual omissions by voters. For instance, consider an election with 100 potential voters, in which Bad gets 30 votes, Good gets 20, and the rest are abstentions. Suppose that all of the abstainers are like Lazy: they would have voted for Good if they had voted, but they abstained. Thus, given how things turned out, Good would have won if 11 of the abstainers had voted. It seems, then, that we want to blame the abstainers for the outcome of the election (as well as the people that voted for Bad). If my arguments in this paper are sound, we should say that none of the individual abstentions was a cause of the outcome. However, we could still argue that the fact that Good didn't get 11 more votes from the pool of abstainers was a cause of the outcome (since, had at least 11 of the abstainers voted, Good would have won). There are different ways that the votes by the abstainers could have gone so that Good got 11 more votes from the abstainers pool. But never mind. If, as I have argued, causation is the vehicle of transmission of responsibility, to the extent that we could make each abstainer responsible for the fact that Good didn't get 11 more votes from the abstainers pool, we could make each abstainer responsible for the outcome of the election. On this view, if a citizen abstains and the outcome of the election is bad, he is responsible for the outcome, not because he causes it, but because he is responsible for something that causes it. If so, the view that causation is the vehicle of responsibility could help us solve the voting problem.<sup>26</sup>

## Notes

\* Special thanks to Randolph Clarke, Juan Comesaña, Ned Hall, Elizabeth Harman, Sarah McGrath, Laurie Paul, Jonathan Schaffer, Judith Thomson, and Stephen Yablo. Thanks also to audiences at the 2003 Syracuse Metaphysics Conference, University of Wisconsin-Madison, University of California-Los Angeles, University of Vermont, and University of Buenos Aires for their helpful comments and criticisms.

1. Henceforth, I will use "responsibility" as short for *moral* responsibility. The thesis that responsibility requires causation is initially very plausible if the responsibility in question is moral, but it is less plausible if it is responsibility of some other type (e.g. legal).
2. For a recent defense of the view that there is no genuine causation by omission, see Dowe (2001). Note that even someone like Dowe, who denies the possibility of causation by omission, feels the pressure to bring the causal and moral realms back in-line to some degree. To this effect, he introduces the term "quasi-causation," and claims that responsibility requires causation *or* quasi-causation.
3. In this paper I will be assuming that agents can be responsible for different categories of things: events or states of affairs in the world, and also their own actions and omissions.
4. See, e.g., Fischer and Ravizza (1998) and van Inwagen (1978).
5. Intuitively, that is. On some views of causation, I caused the explosion by failing to do things that I couldn't possibly have done. In particular, on views according

to which counterfactual dependence is sufficient for causation, I caused the explosion by failing to cast a spell that would magically prevent it (since, had I cast such a spell, the explosion wouldn't have occurred). But recall that Moral Entails Causal says that, if an agent is responsible for an outcome, *it is in virtue of the fact that* some action or omission of his caused the outcome, that is, the fact that some action or omission of his caused the outcome *explains* why the agent is responsible for the outcome. Now, even if my failure to cast the spell were a cause of the explosion, this would not explain why I am responsible for the explosion, since everybody else's failure to cast a similar spell would be a cause too. (Thanks to Jonathan Schaffer for raising this point.)

6. And other people would too. A.M. Honoré writes: "If two huntsmen independently but simultaneously shoot and kill a third person, or two contractors independently fail to deliver essential building supplies on time, it is intuitively clear that each should be held responsible for the death or building delay." (Honoré (2002)) Two Buttons strikes me as analogous to Honoré's two contractors case.
7. See Nagel (1979), ch. 3, and Williams (1981), ch. 2. By contrast, cases of moral *bad* luck are cases where an agent is responsible for some harm partly due to circumstances that are out of his control.
8. This raises the question of how we should distribute the responsibility, which I will not try to answer here. Two Buttons is different from ordinary cases of joint responsibility in that what each of the agents does is *sufficient* (independently of what the other agent does) for the outcome. But, if anything, this provides *more* reason (not less) to believe that each of the agents is responsible in Two Buttons.

Someone might want to try to say that Two Buttons is a case of collective responsibility without individual responsibility: a case where a group is responsible but none of the members of the group is. It has been argued that this phenomenon is possible. I am not convinced. However, even if such phenomenon were possible, Two Buttons is not likely to be an example. The type of case where—some people argue—there is collective responsibility without individual responsibility is one where the behavior of each of the agents is *excusable*. For instance, D. Cooper has argued that a society where some serious injustice has become an everyday practice, and where all of the living members of the society have been brought up to regard it as natural, is a case of collective responsibility without individual responsibility (see Cooper (1972), pp. 88–89). In Two Buttons, however, our individual behaviors *aren't* excusable.

9. I think that there is moral luck of this sort, and I will be assuming that there is in the context of this paper. Notice, however, that it is possible to embrace premise (1) while rejecting the existence of moral luck, by claiming that I am responsible for the explosion both in Two Buttons and in Two Buttons-One Stuck.
10. This is Fischer and Ravizza's "Sharks" case in Fischer and Ravizza (1998). A consequence of the fact that I am responsible for the explosion in Two Buttons is that I can be responsible for outcomes that I couldn't have prevented. (By contrast, Two Buttons-One Stuck and Sharks are cases where I am *not* responsible for an outcome that I couldn't have prevented.)

11. *Pace* D. Parfit, who seems to believe that whether another moral agent is present or whether a natural mechanism is present cannot make a moral difference of the sort I am suggesting. See Parfit (1984), p. 82, and especially fn. 49. Parfit does not say why he thinks this. He might have been led to believe this after noticing that, as I will be claiming, whether another moral agent is present or whether a natural mechanism is present cannot make a causal difference, and then concluding that it cannot make a moral difference. This last step is the step I think we shouldn't take.
12. Is it more implausible to say that every omission overdetermines everything else than to say that every omission causes everything else? This would be so if there were an initial presumption against the existence of widespread overdetermination. There are two main ways in which it has been argued that widespread overdetermination isn't problematic (at least not more than widespread causation). First, if the overdeterminers necessitate each other or are somehow not fully "distinct," such as physical and mental events (this is in reply to the argument against the causal efficacy of the mental by Kim (1998), pp. 44–45). Second, if the overdeterminers are "lawfully yoked," such as the two parts of a rock, which are held together by natural forces (Schaffer (2003)). However, button A's not being depressed and button B's not being depressed are fully distinct conditions and are not lawfully yoked in any way.
13. Could we say that some *other* difference, which comes hand in hand with the difference between a person and a machine, is relevant to my causal powers in these cases? How about the difference in the *certainty* of the outcome? (The thought is: having a person in the other room introduces more uncertainty in what will happen with B than if there is a machine in the room that will inevitably fail.) My reply is twofold. First, in the two rocks case, whether the other rock reaches the window as a result of a mere fluke or as a result of an extremely reliable process is not relevant to whether my rock caused the window to break. All that matters is whether it reached the window, and in what way, not how likely its reaching the window was at the time that it happened. Second, assume that the person in the other room is completely determined to do the wrong thing. This still doesn't get me off the hook. Thus, this strategy won't serve to account for the fact that I am responsible for the explosion in this case. As a result, we would still have to accept my conclusion, or reject a different premise in my argument.
14. A mundane example of causal luck is this: two rocks that, independently of each other, would have been sufficient to break a window collide in midair and lose their momentum; as a result, they don't break the window. There is an important difference between this case and the two buttons cases, though. In this case, the window shattering doesn't occur (or, if it occurs, it has nothing to do with the two rocks acting collectively). In the two buttons, by contrast, the explosion occurs and it has a lot to do with the two button-pushing failures (more on this later).

Although I cannot get into this here, I believe that there are also *action* cases exhibiting a divergence between causal luck and moral luck occurs (see my (ms)). I think there are "mixed" cases too. In particular, some "preemptive preemption" cases might have this feature. For discussion of preemptive preemption cases, see McDermott (1995) and Collins (2000).

15. Recall that I am assuming that omissions can be causes. If so, the causal history of an explosion should contain the failure of the mechanisms that were set to prevent it as well as some "positive" causes.
16. I am oversimplifying, since, presumably, I can clap by moving just one hand. However, we may assume that Jim would also have thought that I was rude if I had just moved one hand. If so, Jim's belief that I was rude seems to have been caused by my failure to simultaneously move my right hand and my left hand in particular ways. Also, one might worry that my failure to clap cannot be fully analyzed in terms of (and, in particular, as the disjunction of) my failure to move my left hand in certain ways and my failure to move my right hand in certain ways because clapping requires a certain *coordination* between the movements of the two hands, not just certain specific movements of each hand. We can avoid this problem by imagining that there is just one way in which I could have moved my hands that could have resulted in my clapping, or by imagining that Jim would still have thought that I was rude if I had moved my hands in any other way.
17. Doesn't this commit us to overly *disjunctive* causes? I think it is no worse (at least, not *much* worse) to believe that  $F(L \wedge R)$  can be a cause than to believe that any omission, e.g.  $F(L)$ , can be a cause.  $F(L \wedge R)$  obtains just in case  $F(L)$  or  $F(R)$  obtains. Similarly,  $F(L)$  obtains just in case I don't move my left hand at all or I move it in either of a number of ways that are not constitutive of clapping. For an argument that there are disjunctive causes, see my (ms).
18. Why don't I just say the following: I am responsible for the explosion in Two Buttons because something I failed to do *entails* a cause of the explosion? Because, as we will see, I am hoping that the new principle will help to account for the moral difference between Two Buttons and Two Buttons-One Stuck. But my failure to depress A also entails a cause of the explosion in Two Buttons-One Stuck.
19. Recall that "responsible" means at least *partly* responsible. What I will argue is that, just as you and I *share* responsibility for the explosion in Two Buttons, we *share* responsibility for our failure to simultaneously depress A and B.
20. As with Moral Entails Causal, the causal connection is also intended to be a *partial* explanation of the fact the agent is responsible for the outcome. Presumably, other conditions are required, such as, roughly, the fact that the agent could foresee that something that he is responsible for would likely cause the outcome.
21. Just as my individual failure to depress A isn't an outcome, our joint failure to simultaneously depress A and B in Two Buttons isn't an outcome. Hence, for all Causal Transmits Moral says, I might be responsible for our joint failure without being responsible for any of its causes.
22. In the jargon of collective action and collective responsibility, it is possible for there to be *individual responsibility* without *individual causation*, as long as there is a specific kind of *collective causation*.
23. For instance, J. Feinberg analyzes "being responsible for a harm" as having acted or failed to act in such a way that (a) one could foresee (or should have foreseen) that the action or omission was likely to lead to the harm, (b) the action or omission caused the harm, and (c) the causal connection between the action or omission and the harm was not deviant. (Feinberg (1970))

24. For an independent argument for this view, see my (forthcoming).
25. Goldman has a peculiar view of agent causation according to which an agent can cause something by omission although omissions themselves are never causes. According to Goldman, Lazy causes Bad's victory, not because his abstention causes it, but because he could have voted for Good and his voting for Good would have been a counteracting causal factor vis-à-vis Good's loss. Never mind. My arguments in this paper are likely to translate into Goldman's view as: Lazy's voting for Good *wouldn't* have been a counteracting causal factor vis-à-vis Good's loss. (As applied to Two Buttons and Two Buttons-One Stuck, the argument goes as follows: had I depressed my button in Two Buttons-One Stuck, my depressing it wouldn't have been a counteracting causal factor vis-à-vis the explosion; but Two Buttons and Two Buttons-One Stuck are causally on a par; thus, had I depressed my button in Two Buttons, my depressing it wouldn't have been a counteracting causal factor vis-à-vis the explosion.)
26. Unlike in the Two Buttons case, in this case each abstainer's omission isn't sufficient for the cause of the outcome. However, the same style of argument that shows that you and I are responsible for the cause of the explosion in Two Buttons seems to apply here. The argument is: someone is responsible for the fact that Good didn't get 11 votes from the abstainers pool; all of them bear equal responsibility for that fact; so, each abstainer is responsible for it.

## References

- Collins, J. (2000) "Preemptive Preemption," *Journal of Philosophy*, pp. 223–34.
- Cooper, D. (1972) "Responsibility and the 'System'," in P.A. French (ed.), *Individual and Collective Responsibility*, Cambridge, MA: Schenkman.
- Dowe, P. (2001) "A Counterfactual Theory of Prevention and 'Causation' by Omission," *Australasian Journal of Philosophy* 79, 2, pp. 216–26.
- Feinberg, J. (1970) "Sua Culpa," in *Doing and Deserving: Essays in the Theory of Responsibility*, Princeton: Princeton U.P., pp. 187–221.
- Fischer, J. M. and M. Ravizza (1998) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge U.P.
- Goldman, A. (1999) "Why Citizens Should Vote: A Causal Responsibility Approach," *Social Philosophy and Policy Foundation* 16, 2, pp. 201–17.
- Honoré, A. M. (2002) "Causation in the Law", in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, Summer 2002 ed., URL: <http://plato.stanford.edu/archives/sum2002/entries/causation-law/>.
- Kim, J. (1998) *Mind in a Physical World*, Cambridge, MA: The M.I.T. Press.
- McDermott, M. (1995) "Redundant Causation," *British Journal for the Philosophy of Science* 46, pp. 523–44.
- Nagel, T. (1979) *Mortal Questions*, New York: Cambridge U.P.
- Parfit, D. (1984) *Reasons and Persons*, New York: Oxford U.P.
- Sartorio, C. (ms) "Disjunctive Causes."
- Sartorio, C. (forthcoming) "A New Asymmetry between Actions and Omissions," *Noûs*.
- Schaffer, J. (2003) "Overdetermining Causes," *Philosophical Studies* 114, 1/2, pp. 23–45.
- van Inwagen, P. (1978) "Ability and Responsibility," *Philosophical Review* 87, pp. 201–24.
- Williams, B. (1981) *Moral Luck*, Cambridge: Cambridge U.P.