The notion of causation is important to moral responsibility since most people think that someone is morally responsible for an event only when that person has caused the event. This is referred to as the entailment claim—that moral responsibility entails causal responsibility (MC). Note that it does not work the other way around: Causal responsibility is (generally) not taken to entail moral responsibility. We can formulate (MC) as follows:

(MC) If an agent A is morally responsible for event e, then A performed an action or an omission that caused e.¹

(MC) is a thesis that seems extremely intuitively plausible. A commitment to (MC) not only has the obvious implications for accounts of moral responsibility and blame/praise, it also has implications for the viability of other philosophical views. For example, fatalists were often criticized for holding views that would undermine the practice of praise and blame. Fatalists believe that our choices make no difference at all to what happens in the world. If one is fated to meet death in Smyrna, then, whatever one chooses to do, that is what will happen. Given the impotence of our choices, we do not cause events to occur.² Thus, we cannot be held responsible for what occurs. And this was seen as a vile, corrupting doctrine leading to moral chaos. However, what underlies the argument is a commitment to (MC). A fatalist who rejected (MC) would not be faced with this problem.

Some philosophers have rejected (MC) due to the metaphysical commitments of the philosophical systems they have adopted. For example, one could read David Hume as rejecting it insofar as one reads him as a causal eliminativist who still believes that we are morally responsible for our actions. Hume’s empiricism seemed to lead him down this road. There are events we call “causes” and events we call “effects.” However, there is
nothing underlying the relation between them in any deep metaphysical sense. Another philosopher who can be read as rejecting (MC) is Gottfried Leibniz. He was not a causal eliminativist, since God, at least, had genuine causal powers, but Leibniz did ascribe to the doctrine of parallelism. God has established a “preexisting harmony” between human wills and events in the world. Thus, when one acts, one does not cause—rather, God has set up a perfect correlation between what is willed and occurrences in the world. Yet Leibniz did not reject the view that we are morally responsible for our actions.

More recently, this entailment claim has been attacked by philosophers who believe it to be false, but not for reasons having to do with the metaphysical commitments of the systems of philosophy they adopt. Rather, they believe that common sense counterexamples can be offered to (MC), ones that don’t depend upon robust empiricism or notions of how God must have configured the universe. First of all, some philosophers hold that omissions are not causes and yet we can still be morally responsible for what we fail to do. They believe that absences or omissions lack power and thus cannot by their very nature produce effects. We might also note that one can “make a difference” and thus be responsible for something, without, strictly speaking, causing it. For example, one could imagine situations in which a judge declared a dead man “innocent” without causing anything to occur in the world. He’s just created an institutional fact that did not exist before. In a similar spirit, some philosophers have discussed so-called “Cambridge” changes—when I leave the room it may be the case that Samantha then becomes the shortest person in the room, but it seems odd indeed to say that I have caused her to be shorter than anyone else in the room (Sosa, 1993).

Another philosopher, John Leslie (1991) holds that we can be responsible for outcomes that we don’t cause but that would not have occurred but for our own actions, via what he terms “quasi-causation.” And, more recently, Carolina Sartorio (2004) also attacks the entailment claim by appealing to disjunctive causation. Both Leslie and Sartorio offer novel cases to motivate their claims, and we will be looking at them later in the paper. What I would like to do in this paper is draw upon findings in social psychology, as well as some of the philosophical literature on causation and moral responsibility, to help develop a strategy to defend (MC) against some of the challenges raised. This is not a strategy to show that (MC) must be true. Rather, it is a strategy that questions the methodology of some of the challenges to (MC). I will return to the counterexamples to (MC) later in the paper and review them in more detail in light of the empirical findings.

When we make attributions of primary causation, that is, when we pick out the cause of an event among a nest of causal factors, it is quite true that we often rely on pragmatic and contextual considerations. Here’s an example of this practice discussed by Hart and Honore (1959).

First, when causal language is used of the provision or failure to provide another with an opportunity, it is implied that this is a deviation from a standard practice or expected procedure; the notions of what is unusual and what is reprehensible by accepted standards both influence the use of causal language in such cases. Hence the case of a house-holder whose prudential storing of firewood in the cellar gave a pyromaniac his opportunity to burn it down would be distinguished from that of the careless friend who left the house unlocked: the fire would not be naturally described as a consequence of the storing of the wood though the loss of the spoons was a consequence of leaving the house unlocked. (p. 56)

This example is one where one figures out who is blameworthy, or responsible, and then assigns causation. We have a sea of causation, and in assigning praise and blame we need to be able to make selections among the respective causal factors. Not all causal factors are blameworthy or praiseworthy, or indicate moral responsibility for the causing agent. This is why (MC) entails only in one direction. The identification in judgment of something’s being the cause depends upon pragmatic factors and may well include moral judgments—Hart and Honore’s cases support this. The householder was not negligent, so his actions were not the cause of the fire, even though it is true that there would have been no fire if he had not been so prudent. Similarly, the careless friend is failing in his responsibilities, so he is identified as the cause, or part of the cause, when the spoons go missing. However, these cases are in no way a threat to (MC) since one, of course, says that there is still a causal connection between what the householder did and the fire. We just don’t attribute causation to him since we don’t view him as morally responsible at all for what happened. Neither case shows we would attribute this moral responsibility in the absence of any actual causal connection. Nor do such cases show that if pressed one would actually deny any causation—rather, the correct thing to say would be something like—“well, of course, the homeowner’s actions were a causal factor, but not the primary cause since he didn’t do anything wrong or unusual.” What we call a “cause” or “causal factor” in many cases depends on what we relegate to the background conditions of the event in question. In the case of the householder, since he behaved normally, we don’t single his actions out. Similarly—in the context as described by Hart and Honore—we wouldn’t say that oxygen caused the fire. But, again, when pressed, we of course acknowledge that it was one of the many causal
Socially undesirable motive
...hide a vial of cocaine he had left out in the open before his parents could see it.

Other cause
Oil spill As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because of some oil that had spilled on the road. As a result, John hit a car that was coming from the other direction.
Tree branch As John came to an intersection, he failed to see a stop sign that was covered by a large tree branch. As a result, John hit a car that was coming in the other direction.
Other car As John came to an intersection, he applied his brakes, but was unable to avoid a car that ran through a stop sign without making any attempt to slow down. As a result, John hit the car that was coming from the other direction.

Consequence of accident
John hit the driver on the driver's side, causing him multiple lacerations, a broken collarbone, and a fractured arm. John was uninjured in the accident.

Complete the following sentence: The primary cause of this accident was **.
(p. 369)

What Alicke discovered is quite interesting. When John's motive is the socially undesirable one—the "culpable" one—he is far more likely to be identified as the primary cause of the accident. As Alicke (1992) himself put it, "With causal necessity, sufficiency, and proximity held constant, the more culpable act was deemed by subjects to have exerted a larger causal influence" (p. 370).

Thus, we have two accidents, let's say, both completely the same except the motive of the driver in question is different. His causal responsibility is deemed much greater when the motive is bad as opposed to when the motive is good. This looks similar to the Hart and Honore thought experiment, but the result is empirically substantiated by looking at the reactions of 174 people.

Further, subjects tended to identify John as a cause more frequently when John was the only agent involved. This was because he "...was viewed as less responsible when another driver contributed to the accident" (Alicke, 1992, p. 370). This factor will be very relevant when we look at the counterexamples again.

Joshua Knobe has recently written a paper in which he is critical of how Alicke models the data. Like Alicke, he believes that moral considerations are crucial to the application of our concept of causation. He argues that "causal attributions are not purely descriptive judgments. Rather, people's willingness to say that a given behavior caused a given outcome depends..."
in part on whether they regard the behavior as morally wrong” (Knobe, 2005a, p. 2). A folk psychological account of these attributions has to address the issue of what work they do for us, “what sort of question a causal attribution is supposed to be answering” (Knobe, 2005a, p. 4). His idea is to use evidence to come to an understanding of our competencies that underlie causal attributions (on analogy with linguistics), and he thinks this is where some normative considerations are to be found.

The sorts of cases he asks us to consider are very similar to the sorts of cases presented by Hart and Honoré—for example, cases where someone falls to live up to a responsibility and thus creates an opportunity for something bad to happen. He also presents a vignette that does not involve omission:

Lauren and Jane work for the same company. They each need to use a computer for work sometimes.

Unfortunately, the computer isn’t very powerful. If two people are logged on at the same time, it usually crashes.

So the company decided to institute an official policy. It declared that Lauren would be the only one permitted to use the computer in the mornings and that Jane would be the only one permitted to use the computer in the afternoons.

As expected, Lauren logged on the computer the next day at 9:00 am.

But Jane decided to disobey the official policy. She also logged on at 9:00 am.

The computer crashed immediately. (Knobe, 2005a, p. 6)

Knobe points out that in this case we seem to think that Jane caused the computer to crash, even though it is also true that if Lauren had not logged on, it would not have crashed. To spell out how these attributions work, Knobe discusses the role of contrast cases; these contrasts are often very subtly signaled in ordinary language—often the speaker relies on features of the context and various inflections and emphases to indicate what the appropriate or intended contrast is. Further, picking out the appropriate contrast is a pragmatic issue, highly contextual. To use a familiar case, if Sally starts a fire that burns down her house, then we would normally attribute the cause of the fire to Sally’s irresponsible behavior, but not to the fact that there is oxygen in the air that fed the fire. This is not to deny that the presence of oxygen in the air is a cause. However, we may more easily say that Sally’s irresponsibility is more of a cause than oxygen because it obtrudes more, or strikes one as more salient. It is not relegated to the background conditions.

Knobe uses the word “obtrude” to indicate that one contrast is more salient or more relevant. He also goes over various considerations people use to determine degrees of salience. However, for our purposes here, I simply want to focus on the normative factors—most specifically, the moral factors. So, when—as in Hart and Honoré’s case—one’s friend fails to lock the door behind him, which he ought to have done, then we attribute, at least in part, the cause of the burglary to him, since the normative consideration of his moral failing has more salience than other causal factors. One of Knobe’s aims is to show that attributions can work somewhat differently than most social scientists believe; they are not merely descriptive but also responsive to normative considerations—like who we think ought to be blamed for something. We attribute causation to Jane in the above case, and not Lauren, because Jane is the one who should be blamed even though both actions were necessary for the computer to crash.

Thus, Knobe does agree with Alicke that normative factors play a crucial role in these attributions—though he disagrees on Alicke’s model for how this works. Alicke’s model moves from (1) the judgment of an agent’s having acted wrongly to (2) the agent’s being blamed to (3) causation’s being attributed to the agent. Knobe (2005a) argues this is not entirely accurate. He believes instead that we move from (1) the judgment of an agent’s having acted wrongly to (2) the attribution of causation to the agent to (3) blaming the agent. We attribute causation prior to the blaming step, at least in some cases—thus, he argues, Alicke’s failure to accurately model the competency. His main case for illustrating this is the following:

George and Harry both work in a large office building. George is the janitor; Harry takes care of the mail.

Every day, George goes through the entire building and empties all of the garbage baskets. Since the building is large, this task normally takes him about one half hour.

One day, George is feeling tired and decides not to take out the garbage.

Harry sees that the garbage hasn’t been taken out. He doesn’t go to take it out himself, since that is not his job.

But it turns out the company is extremely lucky. An accountant had accidentally thrown out an important document, and everyone is overjoyed to find that the trash hadn’t been taken out and hence that the document is still there.

The idea is that we attribute causation to George because it was his failure to act as he ought to have acted that caused the letter to be saved. However, we do not blame him for saving the letter. Thus, attribution of causation must precede the blame step. If this were not the case, then we would need to blame George for saving the letter before holding his failure to be the cause of the letter’s being saved. And we need to be clear about this. Someone might try to argue that George is blameworthy, and we do in fact blame him for his failure to take out the garbage. But
Knobe's point is that even though we do think he did something wrong by not taking the garbage out, and I might blame him for the resulting stinky hallways in the building, I would not blame him for the letter's being saved. Technically, we might note that attribution of causation is independent of the blame step.

Knobe has picked up on the fact that we tend not to blame people for good outcomes, even when the outcomes are the result of something bad that they have done. We tend to blame only for bad outcomes. Someone—let's say a competitor of George's bosses—might blame George for doing something that resulted in the document's being saved, but that's because he would regard the outcome as a bad one.

A critic could respond by claiming that the case has not been accurately described. What George has done wrong is fail to take out the garbage, and we do blame him for that. His saving of the letter wasn't wrong. However, Knobe could respond to this by just slightly changing the case description so that George has spitefully failed to take out the garbage because he thinks the document would be harmful to the company. We still don't blame him for saving the document. Of course, we wouldn't praise him either.

I agree with Knobe that Alicke's analysis doesn't properly accommodate this case. For my purposes I'd like to focus on what they do agree on, which is the fact that normative considerations influence causal attributions—and one of those normative considerations is the consideration that the agent has acted wrongly. However, we might ask whether or not it was that the agent acted wrongly or, rather, somehow "out of the norm"? Consider a slight modification of the George case, one where George does not take out the garbage, though it is his job to do so, but also where George very rarely takes it out even though it is his job. The idea is that this failure to act is normal for George, not unusual at all. On the other hand, Harry's happening to walk down the hall and see the document poking out of the garbage bag was unusual. In this case, I think, we'd probably attribute primary causation to Harry rather than George (another way to put this is that we might say that Harry is more of a cause than George here). This observation is in keeping with some of Joel Feinberg's work on attributing causation. In discussing Bertrand Russell's views on the Cuban Missile Crisis, for example, Feinberg (1970) notes: "Common sense, I would submit, would not necessarily select the last active intervention, in seeking the cause of the crisis, but rather that act or event that was a radical deviation from routine, which was clearly the Soviet construction of missile bases in Cuba" (p. 165, footnote 15). Likewise, he argues that when someone is stabbed to death at a dinner table the cause is not that the table had been laid for dinner with the knife on it. Everybody agrees with this, but Feinberg's explanation is that this is not the cause because "... it is the abnormal deviation that distinguishes the whole incident from other dinners where diners are not killed" (p. 166).

It may be that the hypothesis that judgments of the unusualness of an event or action affecting causal attribution and judgments of normativity affecting causal attribution are not in conflict. However, it may also be that the unusualness hypothesis can account not only for Alicke and Knobe's cases but also nonnormative ones, in which case it would offer a more general explanation, or a more general model. In any case, it is a hypothesis that would be interesting to test more carefully. It would still allow for the contrastivism that Knobe discusses. Whether or not an action or event is "unusual" is relative to a contrast class.

What relevance does the work of Alicke and Knobe have for the sorts of cases we started discussing? Let's look at the counterexamples to (MC) in more detail.

Counterexamples to the Entailment Claim

Again, some philosophers have argued that we need to give up the idea that moral responsibility rests, necessarily, on causal responsibility. For example, some do so out of concerns about being able to spell cut coherently what something like "negative" causation consists in. The idea is that an absence cannot have causal influence. For example, David Armstrong (1999) writes that "Omissions and so forth are not part of the real driving force in nature. Every causal situation develops as it does as a result of the presence of positive factors alone" (p. 177). There are two reasons why some find this view attractive. The first is that many regard causation as a kind of power in the world. Absences are not powers at all—thus the oddity of viewing them as causes. The second reason is a concern for simplicity. Allowing for negative causation will multiply the numbers of causes astronomically. Just think about how many things we are causing to happen right now by our failures to act.

In the moral case, then, if I hold Constance responsible for failing to reach over and flip a switch that would save Bella's life, I am holding her responsible for an omission. It would then follow, if omissions are not actual causes, that she is responsible for Bella's death even though she did not cause it. Thus, we would have moral responsibility without causal responsibility.
Thus, given their concerns about the metaphysical shakiness of negative causation, there are some ethicists who are prepared to give up the view that moral responsibility entails causal responsibility (Harris et al., 1987). Others, such as John Leslie, seem to agree that the entailment claim is false, but for a very different reason, suggesting that we attribute moral responsibility in the absence of causal responsibility when someone does something that, for example, constitutes incredibly strong evidence that the event in question will happen or is happening. What this involves, more specifically, is that subjunctive conditionals track the quasi-effect. He calls this “quasi-causation.” To illustrate, Leslie (1991) asks us to imagine the following fascinating case:

Crossing a plain, I come to what looks like a gigantic mirror. But pushing a hand against it, I feel flesh and not glass. The universe must be symmetric, the flesh that of my double—left-right reversed but otherwise a perfect replica.

The universe must also be fully deterministic, for how else could my double have moved exactly as I did? But never having seen freedom and determinism as incompatible I find this untroubling ...

Do I not actually have twice the power I earlier seemed to have? (p. 73)

Suppose that a woman, Melissa, finds herself in this situation. Of course, there will be Melissa 1 and Melissa 2, a Melissa for each of the universes. Leslie asks us to consider what would happen if one agent picks up a rock and throws it at a bird. The other agent would do the same, given the symmetry. Thus, it is true that were Melissa 1 to throw rock 1 Melissa 2 would throw rock 2, and it is true that were Melissa 2 to throw rock 2 Melissa 1 would throw rock 1. Both Melissa 1 and Melissa 2 would have the feeling of making double the difference. And we get tracking when these sorts of conditionals hold up, so the occurrence doesn’t seem as though it is a freak occurrence. The idea Leslie is employing is this: Even though there is no causal connection between what I do and what my double does, what I do still makes a difference to what she does, and thus my power to bring things about, to make a difference, has doubled. This power relies on the symmetry between these universes. But Leslie goes on to show that symmetry can exist outside of science fiction thought experiments. He notes that two balls may bounce exactly the same height, and that is explainable not by one ball’s affecting the other but by the fact that they are structurally the same and the other background conditions are the same. He terms this not “causation,” but “quasi-causation.” As he notes, fish will respond the same way to the same stimulus, and act identically, though there is no causal interaction between them. There is no causal oomph, so to speak. What we have is a correlation between behaviors that is explainable in terms of structural similarities. In the twin universe cases, the intuition is that Melissa 1’s action can make even more of a difference. This, in turn, might be explainable by the fact that her action generates bad news about how Melissa 2 acts, or by the fact of a (suitably qualified) counterfactual dependence between the two. If this intuition bears scrutiny, then we have another way in which someone could be morally responsible for an outcome even though not causally responsible for it.

Leslie’s case is a stipulation of a Leibniz-like scenario in that there is a perfect correlation between one’s willings and agency and what happens in the world. The parallelism has not been set up by God but exists independently. Given parallelism, had Melissa not thrown the rock, the bird would not have died, in the same way that had Melissa 1 not thrown rock 1, bird 2 would not have died. One might claim that once we get these sorts of robust, tracking, conditional relationships that’s just all there is to causation. However, numerous counterexamples call this claim into question.

Leslie then goes on to apply his findings on quasi-causation to the problem of voting—his view is that if one person fails to vote for the good candidate, then she is responsible, at least partly responsible, for the loss if that candidate loses the election. That is because—in virtue of the symmetry claim—she could have generated the good news but did not; her not voting “meant that” others like her did not vote. Of course, these others share the responsibility for the same reason. Her failure to vote did make an enormous difference even if it lacked the requisite causal power to generate the effect of a loss.

However, another, rather deflationary tack to take is this: There is the illusion of power based on features this case has in common with cases where we ascribe causation to the agent. In the absence of causation, this feeling of power or of “making a difference” is simply an illusion. This is supported by the fact that our intuitions are heavily influenced by the temporal sequence of events—the evidence approach is a counterintuitive basis for “making a difference” when we look at counterfactuals involving past events—for example, either I have spent 1,000 years in hell or not, and if I have, I do not remember it (this is all true). Suppose that the following is true: If I tell a lie to my Mom, then it is likely that I spent a thousand years in hell. Does my telling the lie make a difference—even if “Were I to tell a lie to my Mom, I would have spent 1,000 years in hell already”? Intuitively, not, even though it would be good evidence that I spent 1,000 years in hell, which is a pretty terrible fate.” But perhaps this simply means that we need an evidentialist account with some other constraints.
While Leslie’s case seems to rely on imputing some sort of magical power (he denies this), we can find analogies to the case that Leslie is concerned with in the psychology literature on rationality and game theory. However, this line of thought leads us to adopt an evidentialist decision procedure, which, as already noted, leads to very odd results. In Leslie’s case it seems odd for someone to think he ought to vote because that means that others will vote, in the absence of a causal connection (or, more precisely, his voting constitutes extremely good evidence that others will vote). This seems analogous to opting for the one-box in Newcomb’s problem.

Another writer who has recently argued against the entailment claim is Carolina Sartorio. Her argument rests on a consideration of two cases to be contrasted:

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. (Sartorio, 2004, p. 317)

She will argue that here you are morally responsible for the explosion though you didn’t cause it, by contrasting this case with the following:

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. (Sartorio, 2004, p. 318)

Sartorio claims that the Two Buttons case is a counterexample to the entailment claim. It is a case where one has moral responsibility but the responsible agent has not caused the explosion. The core of her argument rests on our intuitive view that in the Two-Buttons case we think that the agent is morally responsible for not pressing the button, and that in the Two Buttons—One Stuck case we judge lack of causation. Since the omissions are the same in both cases, we must then think that in Two Buttons there is no causation between the failure to depress the button by the single agent and the explosion. Therefore, there is moral responsibility without causation.

Her diagnosis is to suggest the following alternative to the entailment claim: “If an agent is responsible for an outcome, then it is in virtue of the fact that agent is responsible for something that caused the outcome” (Sartorio, 2004, p. 329). One is morally responsible if one causes what in turn causes the outcome in question. Thus, causation is transmissible on her view.

Needless to say, these cases are controversial—Leslie’s in particular. Leslie seems to be arguing for—or perhaps just assuming—evidential decision theory. One ought to do what will generate good evidence that the desired event will occur. Controversial indeed.

Sartorio, on the other hand, is arguing, in effect, that the two acts are causally the same since their intrinsic properties are the same. The failure to press the button is the same in both cases. Thus, if the two acts are the same and one is a cause, then the other is a cause as well. However, the claim they are relevantly the same would be controversial since one could argue that acts are differentiated in terms of relational features as well. Intrinsic features may not be sufficient. She is also, in her analysis, relying on disjunctive causes. These are some very interesting features of her argument which deserve further attention, but I will be focusing on her claim that “I am responsible for the explosion in Two Buttons.” Remember that in this case her action was necessary but not sufficient for the explosion to have been prevented. The other person failed to act, and their conjoined failure to prevent the explosion is what caused it. Thus, if Sartorio is right about her claim, then she does have a case of moral responsibility without causation. However, my claim is that—as in Leslie’s counterexample—we are responding to cues that are associated with causation to assign moral responsibility. We should not deny that she is a cause of the explosion. For example, one factor is that Sartorio has put the case in terms of omissions, which we seem to have some difficulty in regards to causes to begin with (as the quote from Armstrong indicates).

Interestingly, both Leslie and Sartorio apply their analyses to the voting problem. I’ve already mentioned that Leslie views the responsibility to vote as based on generating evidence that others will too—and all those votes will make a difference. Sartorio’s analysis is that if an agent fails to vote for the good candidate, along with a number of other agents who would have voted for the good candidate as well, then she is responsible when
the good candidate fails to win even though she did not cause it—rather, her failure to vote is part of a disjunctive cause for the failure of the good candidate to win. Her omission is a cause of that disjunctive omission, and it is that disjunctive omission that is the cause of the good candidate’s failure to win.

Applying this certainly does not commit one to anything like evidentialism, since one really does view one’s actions in causal terms, but just as part of a disjunctive cause for a particular event. However, one factor that affects Sartorio’s analysis is that we may be influenced by the intuition of irresponsibility when someone fails to vote. This makes it seem like we are precommitted to moral responsibility in these sorts of cases, and it explains why we think Sartorio is right. But this by itself would do nothing to distinguish the case from Leslie’s. For Sartorio then, imagine a case where—for example—a famine caused by drought decimates a population in a far-off land. Suppose also that my driving my car and contributing to global warming was part—a small part—of a huge disjunctive cause leading up to the drought and famine. Am I morally responsible for the deaths? It doesn’t seem so. The larger the disjunction, the less one’s feeling of responsibility. This tracks the lessening contribution to the causal disjunction. But is one morally responsible for the disaster, or morally responsible for simply contributing to the cause of the disaster? Of course, Sartorio is just concerned with presenting a counterexample to the entailment claim, and in doing this she isn’t committed to holding that all cases where one of my actions is part of a disjunctive cause of x are cases where I am morally responsible for x—and I don’t mean to claim that she is doing this. However, what is doing the work in considering these cases—where there is moral responsibility and where there isn’t—seems to depend on preexisting intuitions about moral responsibility. One doesn’t first try to figure out the disjunctive cause issue and then figure out moral responsibility.

What Does the Psychology Research Indicate about How We Fix Moral Responsibility and Causation?

Causation by omission, and thus moral responsibility for failures, can be handled by allowing for negative causation. The previously mentioned objections can be handled by allowing that causes are not powers and by noting that even with positive causation we need to use pragmatic considerations to narrow down the range of relevant alternative causes. Some of these pragmatic considerations, as we’ve discussed, are amply illustrated by some empirical research on how people attribute causation.

However, the other cases will be handled somewhat differently. The Leslie case trades on our intuitions that intending to harm is blameworthy. That is, these sorts of intentions qualify as socially undesirable. Further, even though we have stipulated in the Leslie case that there is no causation at work, these are cases where we have some counterfactual dependence. We are greatly influenced by this connection to feel that there is moral responsibility and, thus, blame is warranted. Even without causation, it is true that but for Melissa 1’s action bird 2 would still be alive. But the psychology of attributions of moral responsibility can explain this misleading intuition. In the case of Melissa 1 we have a perception of very poor motivation and character on her part along with the strong counterfactual connection between her actions and those of Melissa 2, and this makes it more likely for us to attribute to her moral responsibility for bird 2’s death. At least it explains why we have this feeling. The case meets the cues we pick up for attributing both causation and moral responsibility. Even though we have stipulated lack of causation, this cue is still responsible for why we have the feeling of moral responsibility. Of course, relying simply on a counterfactual connection does not stand up to reflection. When we tinker with another cue, like temporal ordering, the feeling of responsibility will diminish. Of course, we could do another study to test whether or not this is so—how much does temporal ordering affect our intuitions about moral responsibility? I suspect, however, that thought experiments are sufficient to show that we will not be willing to say that a is responsible for b when a comes after b.

In Sartorio’s case, again, the fact that the agent has failed to live up to a responsibility affects our intuitions in the way described by Alicke. Sartorio is discussing moral responsibility without causation, and Alicke is discussing our attributions of causation in light of our feelings of moral responsibility. However, Alicke’s research isolates factors that we pick up on in judging, or making attributions of, moral responsibility. He notes that when John acts alone, for example, people tend to judge him as being much more responsible than when it was the case that a nonactor contributed to the accident. The evidence that Alicke has gathered gives us a good indication of how people tend to assign moral responsibility—more agents, less responsibility; bad intentions or motives, more responsibility; and so on. These are the sorts of factors that have a high degree of salience, or, in Knobe’s terminology, “obtrusiveness.”

Much will hinge on how the case is described. The agent in Two Buttons contributes to the disjunctive cause, but so do a lot of other factors. Maybe her mom gave her a particularly interesting magazine to read, for example,
thus making her want to continue reading rather than press the button. Sartorio's analysis in terms of disjunctive causation doesn't show that we aren't just responding first to our view that the agent's failure—itself—is bad and she is morally responsible for that. Very many philosophers have held the view that we have an obligation to vote independent of the expectation that a single vote will decide an election. The thinking is: Yes, my one single vote does not make a difference, and thus my failure to vote does not cause the good candidate to lose, but I have a responsibility to vote nevertheless, and when I fail to vote I am morally responsible for that failure and should be blamed for that. Thus, it is true that one has moral responsibility without causal responsibility in a sense, but it is moral responsibility for failure to vote with respect to lack of causal responsibility for failure to win an election. However, it's not moral responsibility for x without causation for x.

Conclusion

I have had two goals in this paper. The first has been to provide a limited defense of (MC)—limited, because I believe that there will be some odd, Cambridge-change cases that are problematic for (MC). In offering the defense, I have drawn on material in philosophy of law and normative ethics, as well as empirical data from psychological studies by M. D. Alickie—all of which show that many pragmatic factors influence how we make attributions of moral responsibility in situations where a variety of causal factors (or counterfactual ones) obtain.

I have also wanted to show that, though psychological research into these issues of normative significance is very interesting, it should not be seen either as a replacement for, or superior to, traditional "armchair" methods in philosophy. Thought experiments employed by philosophers such as Hart and Honoré, Feinberg, and many, many others have shown that these methods are used to develop extremely nuanced and sophisticated accounts of the sorts of factors that influence our intuitions in normative cases. I have suggested that some factors discussed by Hart and Honoré as well as Feinberg even suggest a more general account of our attributions of causation than the account offered by Alickie and Knobe.

Notes

Earlier versions of this paper were read at Union College, Boston University, and the Conference on Morality and the Brain at Dartmouth College. I would like to thank the members of those audiences for their very helpful comments. Research for this paper was supported by the Leslie Humanities Center at Dartmouth College as part of a workshop on "Morality and the Brain." I would like to thank the Leslie Humanities Center for its support. The paper has also benefited from conversations with Walter Sinnott-Armstrong and Roy Sorensen.

1. Versions of this thesis appear in various places throughout the literature. Sartorio (2004) explicitly discusses a version in her paper. The force of the thesis resides in the fact that it is generally taken to be sufficient for moral innocence with respect to a given event that one did not cause the event in question to occur. Thus, many discussions of moral fault and responsibility assume the thesis.

It is worth noting that the intuition behind the thesis can sometimes be tricky to spell out. For example, we need to note a distinction between someone's being morally responsible for an event and the appropriateness of holding someone morally responsible for an event, which some scholars argue may be done even absent a causal connection between the agent and the event. If someone has a strict instrumental account of the appropriateness of praise and blame, this would be a possibility. An investigation into this wrinkle, however, is beyond the scope of this paper.

2. This depends on treating causes as difference makers.

3. This case is frequently used in the literature. For example, see Hart and Honoré (1959), p. 32.

4. In making this point, Leslie draws on the work of Lewis (1979). There Lewis notes that replicas can provide the basis for great predictive success. "To predict whether I will take my thousand, make a replica of me, put my replica in a replica of my predicament, and see whether my replica takes his thousand" (p. 237).

5. Note that even though it is true that one hand moving would not have happened but for the other, this is not an example of causation even on a counterfactual dependence account simply in virtue of this claim's being true—other conditions are often inserted.

6. Ned Hall (2004) argues that we basically have two fundamentally different concepts of causation. They are (1) what we standardly think of as the production view that holds that causes generate their effects, and then, as well, (2) the counterfactual view that holds that causation can be characterized in terms of counterfactual dependence.

7. I thank Roy Sorensen for this case.

8. Andrew Colman and Michael Bacharach (1997) argue that the "Stackelberg heuristic" in games offers the best payoff strategy, and that this heuristic is "justified by evidentialist reasoning." (p. 1).

9. And Leslie actually does this.