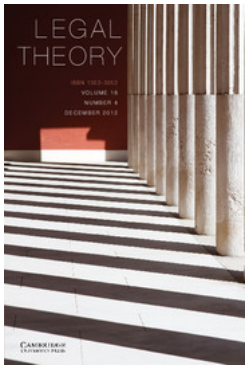
Legal Theory
http://journals.cambridge.org/LEG

Additional services for ***Legal Theory:***

Email alerts: Click here
Subscriptions: Click here
Commercial reprints: Click here
Terms of use : Click here



TWO WRONGS DO NOT MAKE A RIGHT: *RESPONSIBILITY AND OVERDETERMINATION*

Carolina Sartorio

Legal Theory / Volume 18 / Special Issue 04 / December 2012, pp 473 - 490
DOI: 10.1017/S1352325212000122, Published online: 03 July 2012

Link to this article: http://journals.cambridge.org/abstract_S1352325212000122

How to cite this article:
Carolina Sartorio (2012). TWO WRONGS DO NOT MAKE A RIGHT: *RESPONSIBILITY AND OVERDETERMINATION*. Legal Theory, 18, pp 473-490 doi:10.1017/S1352325212000122

Request Permissions : Click here

CAMBRIDGE JOURNALS

Downloaded from http://journals.cambridge.org/LEG, IP address: 68.106.20.197 on 15 Nov 2012

Legal Theory, **18** (2012), 473–490.
© Cambridge University Press 2012 0361-6843/12 $15.00 + 00
doi:10.1017/S1352325212000122

TWO WRONGS DO NOT MAKE A RIGHT:

RESPONSIBILITY AND OVERDETERMINATION

Carolina Sartorio*
University of Arizona

In this paper I critically examine Michael Moore's views about responsibility in overdetermination cases. Moore argues for an asymmetrical view concerning actions and omissions: whereas our actions can make us responsible in overdetermination cases, our omissions cannot. Moore argues for this view on the basis of a causal claim: actions can be causes but omissions cannot. I suggest that we should reject Moore's views about responsibility and overdetermination. I argue, in particular, that our omissions (just like our actions) can make us responsible in overdetermination cases. I go on to provide an account of how this may be possible.

I. INTRODUCTION

In his impressive book *Causation and Responsibility*, Michael Moore examines the relation between metaphysics, morality, and the law.[1] He advocates a picture according to which facts about legal responsibility are grounded in facts about moral responsibility, which in turn are grounded in the relevant metaphysical facts, such as, notably, facts about causation. In this paper, I focus on the relation that Moore sees between the metaphysical facts and the relevant moral facts in a type of scenario usually referred to as "overdetermination." Overdetermination is an intriguing phenomenon that has puzzled legal theorists and philosophers alike. I examine two forms of overdetermination discussed by Moore: overdetermination by the actions of agents or by positive occurrences of some kind and overdetermination by the omissions of agents or by absences of some kind. Moore argues for the existence of a moral asymmetry between these two forms of overdetermination: briefly, agents can be morally responsible for the outcome that ensues in scenarios of the first kind but not in scenarios of the second kind. This moral asymmetry, he argues, is rooted in a causal asymmetry: overdetermining actions

*For helpful comments, thanks to Sara Bernstein, Randolph Clarke, Juan Comesaña, and Marc Johansen.

1. Michael Moore, Causation and Responsibility: An Essay in Law, Morals and Metaphysics (2009).

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

are causes of the ensuing outcomes, but overdetermining omissions are not.

In this paper I challenge the moral asymmetry defended by Moore as well as his claim that the moral asymmetry is grounded in a causal asymmetry. The question about how to assign moral responsibility to agents in scenarios of overdetermination and how to ground those attributions of responsibility in concepts such as causation or other metaphysical concepts raises some interesting challenges, as shown below. I put forth some general principles about responsibility that I think are operative in overdetermination cases and that I hope help to shed some light on this interesting phenomenon.

II. MOORE ON OVERDETERMINATION

As noted above, Moore believes that certain facts about moral responsibility are grounded in the relevant metaphysical facts, in particular, facts about causation. Does this mean that one should hope to settle certain moral questions by first settling the relevant metaphysical questions? As shown below, Moore argues for the soundness of certain "metaphysical-to-moral inferences" that apply to scenarios of overdetermination. I examine the legitimacy of those inferences in due course. But first it will be helpful to start by laying out some more general metaphysical-to-moral inferences embraced by Moore.

In chapter 2 of his book, Moore argues that causation is relevant to moral responsibility in that its presence can increase the blameworthiness of an already blameworthy agent. If, for example, a reckless driver kills a pedestrian as a result of his reckless driving, the fact that his behavior causally resulted in harm renders him more blameworthy than he would have been if he had not caused any harm. In other words, Moore believes in the existence of resultant moral luck: something that is typically beyond the agent's control—namely, whether a certain behavior by an agent causally results in harm—can determine the degree of the agent's blameworthiness.

Now, I am interested in the claim that the agent's responsibility *extends* to an outcome when the agent's behavior actually causes the outcome. (By "extends" I mean that the agent is responsible for the outcome that causally resulted from his behavior in addition to being responsible for the behavior itself, and moreover, he is responsible for the outcome in virtue of being responsible for the behavior that resulted in it.) Arguably, there is a substantial step from this claim to the claim that this results in a genuine increase in the agent's blameworthiness. Some philosophers argue that although the reckless driver who kills someone as a result of his reckless driving is to blame for more things than he would have been responsible for otherwise, he is not more to blame than he would have been otherwise. In other words, what the agent is responsible for, or the scope of

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

his blameworthiness, is different, but how responsible he is, or the degree of his blameworthiness, is the same.[2]

In this way one could grant that causation is relevant to an agent's blameworthiness without committing oneself to the possibility of resultant moral luck concerning the degree of the agent's responsibility. Here I am interested in the claim that causation can result in an extension of an agent's responsibility to outcomes regardless of whether this results in an increase in the agent's blameworthiness. I take it that this is a relatively uncontroversial claim (it is certainly less controversial than the claim about the corresponding increase in the agent's blameworthiness). This claim can be expressed in the form of a pair of metaphysical-to-moral inferences:

M-M 1:

(1.1) If an agent's action for which the agent is responsible caused a given outcome (and other relevant conditions obtained), then the agent's responsibility extends to the outcome.

(1.2) If an agent's action for which the agent is responsible did not cause a given outcome, then the agent's responsibility does not extend to the outcome.

The "other relevant conditions" referred to in (1.1) are the conditions that also have to obtain (in addition to the causal condition) for the agent to be morally responsible for the outcome. Presumably, one such condition is that the agent was able to foresee (or should have been able to foresee) that his action would likely cause the outcome. But there could be others.

In Moore's view, the M-M 1 inferences do not apply to agents' omissions. This is because Moore believes that omissions cannot be causes. (He believes this because he believes that omissions are mere absences and he believes, in turn, that absences cannot be causes. See, especially, chapter 18.) Still, Moore thinks that some omissions can make us morally responsible for outcomes such as harms. For example, a lifeguard can be responsible for a swimmer's death in virtue of having omitted to jump into the water to try to save him. Hence the absence of causation does not entail the absence of responsibility in the case of agents' omissions.

Our omissions can make us morally responsible for harms, in Moore's view, because there can be *counterfactual dependence* between our omissions and those harms (see, again, chapter 18). Counterfactual dependence is the relation that obtains between X and Y whenever, had X not occurred, Y would not have occurred either. Thus an outcome counterfactually depends on what an agent omits to do (say, the agent's omitting to perform an act A) just in case, had the agent failed to omit to A (i.e., had he performed act A), the outcome would not have occurred. For example, a swimmer's death counterfactually depends on the lifeguard's failure to attempt a rescue when, had the lifeguard attempted a rescue, he would have saved the

2. *See, e.g.*, J.J. Thomson, *Morality and Bad Luck*, 20 METAPHILOSOPHY 203–221 (1989); *reprinted in* MORAL LUCK 195–215 (D. Statman ed., 1993); and M. Zimmerman, *Taking Luck Seriously*, 99 J. PHIL. 553–576 (2002).

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

swimmer. The lifeguard can be responsible for the swimmer's death to the extent that the death counterfactually depended on his omission to attempt a rescue.

Causation and counterfactual dependence are, in Moore's view, the only two metaphysical bases for (noninchoate)[3] responsibility, that is to say, they are the only two metaphysical relations that can ground an agent's moral responsibility for a given outcome in the world. Hence, in the case of omissions, Moore would endorse the following pair of metaphysical-to-moral inferences:

M-M 2:

(2.1) If a given outcome counterfactually depended on an agent's omission for which the agent is responsible (and other relevant conditions obtained), then the agent's responsibility extends to the outcome.

(2.2) If a given outcome did not counterfactually depend on an agent's omission for which the agent is responsible, then the agent's responsibility does not extend to the outcome.

The second inference in M-M 2, claim (2.2), has important implications in cases of overdetermination involving omissions. Overdetermination is typically defined in the following way: an outcome is overdetermined by two events if the outcome would not have occurred in the absence of both events but each event was independently sufficient for the outcome's occurrence, that is, the outcome would still have occurred if either of those events had occurred without the other.[4] More particularly, our focus here is on a specific kind of overdetermination sometimes called *symmetrical* overdetermination. There is symmetrical overdetermination when the overdeterminers' contributions are on a par.[5] A classical example of symmetrical overdetermination is two events of shooting a victim through the heart at exactly the same time. Symmetrical overdetermination is contrasted with another kind of overdetermination, called "asymmetrical" or "preemptive" overdetermination, in which there is an asymmetry between the contributions of the overdeterminers; one of them, the preempting event, is a cause, and the other one, the preempted event, is not. Following common usage, here I use "overdetermination" to mean symmetrical overdetermination.

As we see in Section IV, this definition of overdetermination is not fully accurate; in particular, the counterfactual account of the relevant notion of sufficiency is flawed. But it is at least a rough approximation to the phenomenon of overdetermination, and it will have to do for the time

3. Noninchoate responsibility is opposed to responsibility for mere attempts.

4. D. LEWIS, *Causation*, *in* 2 PHILOSOPHICAL PAPERS 159–213 (1986), at postscript E.

5. Alternatively, some people say that there is symmetrical overdetermination when the overdeterminers are both *causes*. Lewis does not go this route because he believes that overdeterminers do not cause the subsequent effects (taken individually, that is; they do "collectively," since the mereological sum of the overdeterminers causes the effect). Here I am working with Lewis's formulation, for obvious reasons (Moore believes that omissions are never causes, so omissions could never give rise to overdetermination on this alternative view).

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

being. Note that the definition focuses on overdetermination by two events. Overdetermination by two absences, say, two omissions, would be defined in a similar way: an outcome is overdetermined by two omissions when it would not have occurred in the absence of both omissions and each omission was independently sufficient for its occurrence. (Perhaps one should add a symmetry condition here, too, stating that the contributions of the two omissions are on a par. Note, however, that in this case one could not understand this condition as making reference to the *causal* contributions of the omissions since, on a view like Moore's, omissions are never causally efficacious. Hence one would have to find an alternative way of drawing the distinction between the symmetrical cases and the asymmetrical cases. Since here I am concerned with only the symmetrical cases, I basically sidestep this delicate issue in what follows.)[6]

An example of overdetermination by two omissions is the following scenario. Two buttons have to be depressed at a given time in order to prevent an explosion. There is one person in charge of each of the buttons. However, when the time comes, both of the persons in charge fail to depress their buttons, and the explosion ensues as a result. In this case the explosion is overdetermined by the two agents' omissions. I call this scenario "Two Buttons."[7]

Note that a main feature of overdetermination scenarios is the absence of counterfactual dependence between each of the individual overdeterminers and the outcome. If two acts of shooting a victim through the heart overdetermine the victim's death, the death does not counterfactually depend on either of the individual overdeterminers (it would still have occurred if one of the shootings had occurred without the other). Similarly, in Two Buttons the explosion does not counterfactually depend on the agents' individual failures to depress their buttons (it would still have occurred if one failure had occurred without the other). Hence, even if the other conditions for responsibility are met (say, both agents had reason to believe that they could prevent the explosion by depressing their buttons, they did not do it because they wanted the explosion to happen, etc.), it follows from claim (2.2) above that neither agent is morally responsible for the explosion in Two Buttons.

This result does not extend to scenarios of overdetermination involving actions. For Moore believes that any overdetermining positive occurrences are causes of the ensuing outcome. Hence M-M 1 (in particular, claim (1.1)) entails that agents are morally responsible for the outcome in virtue of their overdetermining actions, in the right circumstances. For example, if depressing one button is enough to send an electrical shock that will kill a person, and two agents depress their buttons simultaneously, (1.1) entails

6. I revisit this issue briefly at the end of the paper. *See* Section V *infra*.

7. The example is taken from C. Sartorio, *How to Be Responsible for Something without Causing It*, 18 PHIL. PERSP. 315–336 (2004).

CAMBRIDGE JOURNALS

http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

that in the right circumstances, both agents are responsible for the victim's death. This results in an asymmetry between scenarios of overdetermination involving actions and those involving omissions, an asymmetry that concerns the moral responsibility of the agents involved in those scenarios, and one that is rooted in an asymmetry between the causal powers of actions and omissions.[8]

III. TWO WRONGS DO NOT MAKE A RIGHT—PART I

I argue that we should reject the moral asymmetry advocated by Moore between the two forms of overdetermination. In particular, we should reject Moore's claim that agents in overdetermination scenarios of omission are not responsible for the overdetermined outcome. I proceed as follows. First, in this section, I argue that the claim that agents are not responsible for the outcome in overdetermination scenarios involving omissions is extremely counterintuitive, and I pinpoint, in rough terms, the source of the counterintuitiveness. Then, in the next section, I attempt to make more precise the nature of that source.

Imagine that they tell you that depressing the button located right in front of you will prevent an explosion, but you fail to depress the button because you want the explosion to happen. Imagine that depressing the button would indeed have prevented the explosion. Clearly you are responsible for the explosion in this case. Now imagine that, unbeknownst to you, there is a second agent who is told the same thing as you (he is told that he could prevent the explosion by depressing the button located right in front of him) and imagine that he also fails to depress his button because he wants the explosion to happen. Imagine, as in the Two Buttons scenario described above, that both buttons had to be depressed in order for the explosion to be prevented, so the explosion was overdetermined by the two omissions. Could the existence of this other agent and this other button relieve you of responsibility for the explosion? Intuitively, no. For if it did, then *your* existence and the existence of your button would also relieve *him* of responsibility, and then nobody would be responsible for the explosion in that case. However, how could the fact that more people failed to do what they were supposed to do result in no one's being responsible for the explosion? Two wrongs do not make a right.[9]

8. Moore discusses scenarios of symmetrically culpable co-omitters in MOORE, *supra* note 1, chs. 5, 6, and 18 (*see*, especially, *id.* at 450). Moore also discusses scenarios of asymmetrical culpability involving omissions and other absences (in particular, see his discussion of the famous "thirsty traveler" case, *id.* at 466–467). Although I believe that the main reason to reject Moore's view in overdetermination scenarios also applies to the asymmetrical cases, for ease of exposition I am concerned only with cases of symmetrical culpability. Asymmetrical cases give rise to a range of puzzling new questions, which I prefer to bypass in this paper. (*See* n. 28 *infra*.)

9. Moore recognizes that many people find it counterintuitive to say that agents are not responsible for the outcome in cases like this, but he thinks that such an intuition can be

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

Here is a real-life example with a similar structure. Every time there is an election in the United States, many people who could have voted in the election fail to vote. The outcome of the election is always overdetermined, and almost always overdetermined by a large margin. Still, it is often the case that if several of the people who failed to vote had voted for a certain candidate (a losing candidate), that candidate would have won the election. Just as we want to hold people who voted in an election responsible for the outcome, we want to hold people who did not vote responsible for the outcome. But, given that the outcome is (in all realistic cases) overdetermined, it follows from M-M 2 (in particular, claim (2.2)) that none of the people who failed to vote is responsible for the outcome of the election (although many of the people who did vote are presumably responsible for the outcome).

The election case is particularly interesting because, given that the outcome is usually overdetermined by such a large margin, the agents involved do not even have good reason to believe that they can make a difference to the outcome. (When the advertisements for a certain candidate try to convince you that you can "make a difference" by voting for a certain candidate, they are usually simply false or only metaphorically true. In other cases they are true but irrelevant, as when they read: "*Together* we can make a difference!") Still, we think that people can be responsible for the outcome of an election by voting or by failing to vote. In the Two Buttons case I stipulate that both agents believe (falsely) that they can make a difference, since it is particularly clear that we would want to blame them for the outcome then. But, as the election case suggests, believing that you can make a difference might not actually be required for responsibility in every case of this kind.[10]

At any rate, it seems clear that the claim that agents are not responsible in overdetermination scenarios of omission is very counterintuitive and that its counterintuitiveness can be captured by means of the slogan "two wrongs do not make a right" (hereafter, "The Slogan"). By "wrongs," of course, I mean wrongs that are blameworthy. The Slogan claims that two blameworthy wrongs cannot together constitute a blameless right (the mere fact that someone else acted wrongly and in a way that deserves blame does not get a blameworthy wrongdoer off the hook).[11] But this is obviously very rough; in particular, as we see next, there are some ways in which two

explained away as an illicit kind of "moral clumping" phenomenon (*see* MOORE, *supra* note 1, at 450). I argue that on the contrary, the intuition can be adequately justified.

10. The election case gives rise to a puzzle, because it is unclear how we can assign responsibility to the agents if they have every reason to think that their individual vote cannot make a difference. But it seems clear that we want to hold agents responsible in scenarios of that kind at least in some cases; the problem is: How?

11. Here I am not concerned with other forms of responsibility, such as praiseworthiness, since they are not the focus of Moore's work. It seems to me clear, however, that similar considerations apply to praiseworthiness. Although it is harder to formulate The Slogan for praiseworthy acts ("two rights do not make a wrong" plainly does not capture it), a similar thought holds: your act cannot cease to be praiseworthy just because others act in similarly praiseworthy ways.

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

wrongs *can* make a right. So in what follows I proceed to clarify the sense in which The Slogan states something true and I explain how The Slogan, understood in that way, applies to overdetermination cases, including cases of overdetermination involving omissions. This takes us through a brief detour into the nature of overdetermination.

IV. TWO WRONGS DO NOT MAKE A RIGHT—PART II

Imagine two shooters who simultaneously try to kill a victim but whose bullets collide in mid-flight and are deflected away from the victim. Call this case "Colliding Bullets." Clearly, the two shooters are not responsible for the victim's death in this case. The simplest scenario of this kind is one where the victim does not even die. But note that the two shooters would still escape responsibility for the death even if the death were to happen in some other way—say, if a third person had shot him a few seconds later. The reason the two shooters escape responsibility for the death in these cases is that their contributions interfere with each other in such a way that they are no longer sufficient for the death. If the death still happens, it is because a different causal process resulted in the death. Plainly, two wrongs *can* make a right in this way.

Now, note that our target scenarios, overdetermination scenarios, are importantly different. Overdeterminers do not interfere with each other or cancel each other out in the same way they do in Colliding Bullets. No additional causal process is needed to make the outcome happen, since each overdeterminer is still sufficient for the outcome's occurrence, even in the presence of the other overdeterminer. On reflection, it seems unproblematic to suggest that two wrongs can make a right if, by virtue of the two wrongs happening together, they are not actually sufficient for the harm. However, it does seem problematic to suggest that two wrongs make a right when each of the two wrongs is actually sufficient for the outcome, even in the company of the other.

The claim that I want to defend, then, is not the claim that two wrongs can never make a right, but the claim that two wrongs cannot make a right unless certain special circumstances obtain. These special circumstances are ones in which two conditions that would have been sufficient for the occurrence of an outcome if they had occurred in isolation from each other are not in fact sufficient for the outcome when they occur simultaneously. Note also that my main reason for wanting to say that two wrongs cannot make a right (unless special circumstances obtain) is not that an agent's responsibility cannot depend on what others do or on factors that are beyond the agent's control. In accepting that two wrongs can make a right in some (special) circumstances, I am granting that sometimes an agent's responsibility can be partly a matter of luck (for whether your bullet collides with another

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

person's bullet is not fully within your control).[12] The claim I am making is that two wrongs cannot make a right unless luck intervenes in such a way that the individual contributions are not actually sufficient in the relevant way for the harm to occur.

Recall that when I give the standard definition of overdetermination in Section II, I state that it could only be taken as a first approximation to the true nature of the phenomenon. We can now see why. Schematically, the definition is the following:

(OD) Two actual events C and D overdetermine another actual event E iff:
(i) If neither C nor D had occurred, then E would not have occurred.
(ii) If C had occurred without D, or if D had occurred without C, then E would have occurred.
(iii) C and D were on a par (their contributions to E were symmetrical).

Scenarios such as Colliding Bullets can be used (with a minor twist) to show that OD is flawed. Imagine, again, that after the two shooters' bullets collide, the victim's death still occurs due to the existence of a third shooter. Imagine also (this is the minor twist on the example) that the third shooter would not have been motivated to shoot the victim in any other case, in particular, he would not have shot the victim if neither shooter had shot at him. Call this case "Colliding Bullets 1." OD entails that the death is overdetermined by the two shooters in this case. For (given the minor twist) the death would not have occurred if the two shootings had not taken place, it would have occurred if only one of the shootings had occurred, and there is no asymmetry between the contributions of the two shootings. But, clearly, the two shootings do not overdetermine the death in this case.[13]

Consider, also, this other scenario. Imagine that this time there is no third shooter; after the two shooters' bullets collide with each other, they reach the victim and kill him. However, imagine that the two bullets are slowed down considerably by the collision and, as a result, by the time they reach the victim, both bullets are needed to kill him. Call this scenario "Colliding Bullets 2." Again, this is not a case of overdetermination. However, OD entails that the death is overdetermined by the two shooters in this case, too. For, again, the death would not have occurred in the absence of both shootings, it would have occurred if only one of the shootings had taken place, and there is no asymmetry between the contributions of the two

12. Although recall that (as I point out in Section II) there might be a further step from there to the claim that this is genuine moral luck about the *degree* of the agent's blameworthiness.

13. Note that, given that in this case the third shooter shoots only because the two shooters shoot and miss, we might want to say that the two shootings still contribute to the death (they are contributing causes, even though they are not overdeterminers). If so, Colliding Bullets 1 is not a scenario where two wrongs make a right but one where two wrongs make something like a "lesser wrong." The same goes for Colliding Bullets 2, the next case discussed in the text.

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

shootings. The two Colliding Bullets scenarios, then, show that OD fails to offer sufficient conditions for the existence of overdetermination.[14]

What has gone wrong? What these examples suggest is that the key notion of sufficiency involved in overdetermination cases cannot be spelled out in simple counterfactual terms as the claim that the occurrence of one of the overdeterminers in the absence of the other would have resulted in the outcome's occurrence. The colliding bullets are sufficient for the victim's death in this sense but they are not sufficient *in the sense that matters for overdetermination*.

What is the right account of overdetermination, then? It is hard to say, and this is not the place to try to give a better account.[15] Instead of attempting to do this, I will simply point at one possible direction in which one could go. It seems to me that one could try a more sophisticated counterfactual account that appeals to strategies sometimes used to give an account of the concept of causation itself. In particular, one could appeal to the strategy of uncovering certain dependencies that exist while holding certain actual facts fixed. This is a strategy employed by, for example, Stephen Yablo[16] in giving his counterfactual account of causation. Yablo argues that although sometimes effects do not counterfactually depend on their causes (as in cases of preemption), in those cases there is still "de facto" dependence, or dependence *in the actual circumstances*: the effect depends on the cause holding fixed certain facts about potential routes to the effect (facts in virtue of which those other routes are closed or causally inefficacious).

Similarly, one could argue that the relevant notion of sufficiency behind overdetermination is a kind of *de facto sufficiency*. Take the two bullets in Colliding Bullets 2. If either one of those two bullets had not been shot, the victim would still have died. However, if either one of those two bullets had not been shot *but somehow the other bullet had retained its actual momentum*, then the victim would not have died. In other words, neither bullet is sufficient for the victim's death when one holds fixed its actual momentum. Similarly for Colliding Bullets 1: if either bullet had not been shot, then, holding fixed the fact that the other bullet was deflected away from its path, the death would

14. Arguably, OD also fails to offer necessary conditions. Imagine a "shy" bullet that can move only if it senses other bullets moving around it. It seems that an ordinary bullet and a shy bullet can overdetermine a death when shot simultaneously, although had the shy bullet been shot without the ordinary bullet, the death would not have occurred (condition (ii) fails). Thanks to Sara Bernstein for the example.

15. N. Hall, *The Intrinsic Character of Causation*, *in* 1 OXFORD STUDIES IN METAPHYSICS 255–300 (D. Zimmerman ed., 2004), offers a different account that seems to avoid the problems raised by the colliding bullets scenario. On Hall's view, an overdetermination scenario is, roughly, a scenario that can be broken up into two or more intrinsic causal structures culminating in the effect. But it is unclear that this account will capture every case of overdetermination. For instance, we could imagine a Colliding Bullets scenario in which each bullet contributes to the intrinsic properties of the causal process involving the other bullet (say, it helps to determine the other bullet's momentum after the two bullets collide) but where the resulting momentum of each bullet is still enough to kill the victim. This is a case of overdetermination but it would be hard to capture it by appeal to Hall's account.

16. S. Yablo, *De Facto Dependence*, 99 J. PHIL. 130–148 (2002).

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

not have occurred. By contrast, in a genuine scenario of overdetermination, such as one where the two bullets pierce the victim's heart at exactly the same time without colliding with each other, the shooting of each bullet is sufficient for the death in the relevant sense (if either bullet had not been shot, then, holding fixed the actual momentum of the other bullet, the death would still have occurred).[17]

Note that an account of overdetermination along these lines (a "de facto" counterfactual account of overdetermination) is still in some important ways a counterfactual account, but it is a counterfactual account where the actual facts play an important role, too (and not just by virtue of fixing the truth of the relevant counterfactuals). Setting aside the question of whether an account of this type could succeed, it seems to me that this is a main virtue of such an account. For an account of this type strives to capture an important insight about overdetermination brought out by the Colliding Bullets scenarios: the fact that overdetermination has to do in some important sense with the *actual* contributions of the overdeterminers. We feel that two colliding bullets cannot be overdeterminers if the *actual* momentum of each bullet is not sufficient to bring about the death by itself, even if each one of them *would* have been sufficient to bring about the death by itself. "Actual sufficiency", not "counterfactual sufficiency", is what seems to matter for overdetermination.

Now we may finally return to The Slogan ("two wrongs do not make a right"). My suggestion is that we should read it as the claim that two wrongs do not make a right when each of the two wrongs is sufficient for the occurrence of the ensuing harm, in the sense of sufficiency that matters. That is to say, two wrongs do not make a right when each of the two wrongs is sufficient for the occurrence of the harm not in the sense of counterfactual sufficiency but in the sense of *actual* sufficiency.

What follows from this? One thing that follows is that in the case of two blameworthy acts, where the actual contribution of each one of the two acts is sufficient for the occurrence of a harm, the two agents are responsible for the harm in virtue of their blameworthy acts. The different shooting scenarios examined in this section involve acts by two agents. But it is possible to apply The Slogan, thus understood, to omissions as well. In the case of omissions, it is a bit harder to get a grip on the notion of actual sufficiency and the difference between actual sufficiency and counterfactual sufficiency. But the difference still seems to exist. Consider the following variant on the Two Buttons scenario discussed above. Again, if two buttons were to be depressed simultaneously, an explosion would be prevented. Again, each of the agents in charge intentionally fails to depress his button. However, imagine that, unbeknownst to them, an emergency mechanism

17. Note that an account along these lines would also have the right result in the shy bullet case described in n. 14 (the scenario that seems to show that OD fails to offer necessary conditions for overdetermination).

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

has been set up to go off automatically if (and only if) *both* buttons fail to be depressed. So when the two agents fail to depress their buttons, the emergency mechanism goes off and prevents the explosion. This is another scenario where two wrongs make a right.

And we can see the distinction between counterfactual and actual sufficiency here too. Although in this case each omission is counterfactually sufficient for the explosion (if one agent had omitted to depress his button but the other agent had not, then the explosion would not have been prevented), neither omission is *actually* sufficient for the explosion. By contrast, in the original Two Buttons scenario, the two omissions *are* actually sufficient for the explosion. Hence, whereas two wrongs do make a right in the scenario where the emergency mechanism goes off and prevents the explosion, they do not in Two Buttons.

I conclude that the right understanding of The Slogan undermines Moore's asymmetry between overdetermination scenarios of action and overdetermination scenarios of omission. For in scenarios of both kinds, as in all genuine cases of overdetermination, the two wrongs are each actually sufficient for the harm. And whereas two wrongs can make a right when they are each merely counterfactually sufficient for the harm, they cannot make a right when they are each actually sufficient for the harm.[18]

V. IN SEARCH OF A RESPONSIBILITY BASIS

Recall the metaphysical-to-moral inferences that Moore embraces in the case of omissions:

M-M 2:

(2.1) If a given outcome counterfactually depended on an agent's omission for which the agent is responsible (and other relevant conditions obtained), then the agent's responsibility extends to the outcome.

18. There are some scenarios where it might appear, at first sight, that two *actually* sufficient wrongs make a right. But I think that these are cases where we do not have two wrongs to start with. Consider, for example, this variant on the Two Buttons case. The two agents, X and Y, find out about each other's existence. X is told by a reliable source that Y will not press his button, and Y is told by another reliable source that X will not press his button. As a matter of fact, however, both want the explosion to happen, so they fail to press their buttons wishing it to happen. Arguably, neither X nor Y is responsible for the explosion in this case, but their omissions are actually sufficient for the explosion. However, I think that to the extent that we think the agents are not responsible for the explosion in this case, it is because we think their omissions are not blameworthy. Failing to do something that you have very good reason to believe will not do any good is not usually blameworthy, regardless of what your intentions regarding the outcome are. (If a child is drowning, and a bystander has good reason to believe that throwing him a lifesaver will not be enough to save him, then the bystander is not blameworthy for the child's death in virtue of not throwing him the lifesaver, even if he wanted the child to die.)

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

(2.2) If a given outcome did not counterfactually depend on an agent's omission for which the agent is responsible, then the agent's responsibility does not extend to the outcome.

I argue above that we should reject M-M 2. In particular, I argue that we should reject claim (2.2): when presented with cases of overdetermination by two omissions such as the Two Buttons case (where the outcome fails to depend counterfactually on the individual omissions), we should not conclude that the agents involved are not responsible for the outcome. This means that if our initial ideas about metaphysical bases for responsibility conflict with the claim that the agents are responsible for the outcome in these cases, we should revise not the claim about the agents' responsibility but instead our initial ideas about what constitutes an appropriate metaphysical basis for responsibility. In essence, this is what Moore thinks we should do in paradigm cases of responsibility by omission (where, he believes, causation fails to ground responsibility): we should take as a starting point the assumption that the agent is responsible and then look for an appropriate relation (such as counterfactual dependence) to ground his responsibility. What I am suggesting is that here, too, we should reason in this way: we should take the assumption that the agents are responsible as a starting point and then look for an appropriate way to ground their responsibility.

Now, what could this basis for responsibility be? If one believed, unlike Moore, that causation by omission is possible, then one could of course try to argue that the relevant responsibility basis is causation. That is to say, one could try to claim that the agents in Two Buttons cause the explosion, although the explosion is not counterfactually dependent on their individual omissions. Now, I do not think that this is a promising strategy. For even if causation by omission were possible, I think that it would be very implausible to suggest that the agents in Two Buttons cause the explosion. I argue for this elsewhere,[19] so here I am limiting myself to a brief outline of the main argument. It is the following.

Imagine a variant on the case, "Two Buttons–One Stuck," where one button is controlled by a person and the other button is controlled by an automated mechanism. On this particular occasion, where both buttons have to be depressed to prevent the explosion, the agent fails to depress his button, and the mechanism also fails (the button controlled by the mechanism becomes stuck). Presumably, the agent's omission is not a cause of the explosion in this case. But, I argue, there is no metaphysically significant difference between Two Buttons and Two Buttons–One Stuck in virtue of which the agent could be a cause of the explosion in Two Buttons and not in Two Buttons–One Stuck. Therefore the agent is not a cause of the explosion in Two Buttons. In my opinion, this is what makes Two Buttons (and in

19. Sartorio, *supra* note 7.

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

general, overdetermination cases involving omissions) particularly interesting: the fact that in these cases, agents seem to be morally responsible for the ensuing outcomes without causing those outcomes.[20]

If not causation, what could the responsibility basis be in cases like Two Buttons? In my *Philosophical Perspectives* article, I give a partial answer to this question.[21] The first thing to note is that if we allow causation by absences, then, although neither of the individual failures is a cause of the explosion in Two Buttons, the failure of the two buttons to be depressed simultaneously presumably is. (If one does not believe in causation by absences, then one can simply rephrase this in terms of counterfactual dependence: although the explosion does not counterfactually depend on the individual failures, it does depend on the failure of the two buttons to be depressed simultaneously. One can then run the rest of the story in terms of counterfactual dependence.) Note that the fact that agent X fails to depress his button entails the fact that the two buttons are not depressed at the relevant time, since the fact that the two buttons are not depressed at the relevant time obtains just in case *either* X does not depress his button *or* Y does not depress his (or neither does). The same goes for the fact that agent Y fails to depress his button. In other words, the fact that the two buttons are not depressed at the relevant time is a disjunctive fact, which obtains just in case at least one of the individual failures obtains. I will refer to this fact as "D."

Should we say, then, that each of the agents in Two Buttons is responsible for the explosion because the fact that he omits to depress his button *entails* D, and D is a cause of the explosion (or is something on which the explosion counterfactually depends)? Although it is quite natural to want to say this, again, unfortunately I do not think it works. For all of this is true of the agent in Two Buttons–One Stuck, where we (presumably) do not want to say that the agent is responsible for the explosion.[22] What we should say, instead, is that the agents in Two Buttons are responsible for the explosion because they are responsible for D, which in turn causes the explosion (or is something on which the explosion depends). In contrast, the agent in Two Buttons–One Stuck is *not* responsible for D. Hence, even if the fact that he fails to depress his button entails D, and even if D is a cause of the explosion (or something on which the explosion depends), the agent does not come out as responsible for the explosion in Two Buttons–One Stuck.

20. By the way, I sympathize with Moore's claim that overdetermining *actions* are causes. However, if this turned out to be wrong and overdeterminers were never causes (as, for example, LEWIS, *supra* note 4, believes), I would say the same thing about scenarios of overdetermination involving actions: agents are responsible for the outcome in these cases, and their responsibility is grounded in some other way.

21. Sartorio, *supra* note 7.

22. Moore agrees with me about the agent's lack of responsibility in cases like Two Buttons–One Stuck. He then goes on to suggest that in light of this fact, it is very hard to justify the claim that the agents are responsible in cases like Two Buttons (*see*, in particular, MOORE, *supra* note 1, at 451 and 467). In what follows, I suggest a way in which to ground the difference in responsibility between the two scenarios.

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

That is what I say in my *Philosophical Perspectives* article.[23] Obviously, a full account of the agents' responsibility in Two Buttons would have to say something about *why* the two agents in Two Buttons are responsible for D but the one agent in Two Buttons–One Stuck is not. I did not attempt to do that back then because my aim at the time was not to give an account of the bases for responsibility but only to explain how one can *consistently* claim that the two agents are responsible in Two Buttons but the one agent is not responsible in Two Buttons–One Stuck. The thought was that, just as it is plausible to think that the agents in Two Buttons are responsible for the explosion but the agent in Two Buttons–One Stuck is not, it is also plausible to think that the agents in Two Buttons are responsible for D but the agent in Two Buttons–One Stuck is not.

Now, if one wanted to do more than this, what could one try to say? How could one try to explain the difference in the agents' responsibility between Two Buttons and Two Buttons–One Stuck with respect to D? This is what I want to explore in this final part of the paper.

As I explain above, D is a disjunctive fact: it is the fact that obtains just in case agent X does not depress his button (at the relevant time, t) or agent Y does not depress his button (at t). X's failure to depress his button entails D, and so does Y's failure to depress his button. So consider the question: Under what conditions is it plausible to claim that an agent is or is not responsible for a disjunctive fact? In the literature on moral responsibility, several people endorse a principle along the following lines (I call it the "Responsibility for Disjunctions" [RD] principle):

> (RD) If both A and B obtain, and an agent is responsible for A but not for B, then the agent is not responsible for the disjunctive fact: A or B.[24]

Obviously, RD is not a general principle about responsibility for disjunctive facts, since it merely states a sufficient condition for the absence of responsibility for such facts. Still, it is helpful for our purposes here.

23. Sartorio, *supra* note 7.

24. *See* P. van Inwagen, *Ability and Responsibility*, 83 PHIL. REV. 201–224 (1978), sec. III; R. Heinaman, *Incompatibilism without Alternative Possibilities*, 64 AUSTRALASIAN J. PHIL. 266–276 (1986); *reprinted in* PERSPECTIVES ON MORAL RESPONSIBILITY 296–309 (J. Fischer & M. Ravizza eds., 1993), at 306; W. Rowe, *Causing and Being Responsible for What Is Inevitable*, 26 AM. PHIL. Q. 153–159 (1989); *reprinted in* PERSPECTIVES ON MORAL RESPONSIBILITY 310–321 (J. Fischer & M. Ravizza eds., 1993), at 320; and J. Fischer & M. Ravizza, *Responsibility for Consequences*, *in* PERSPECTIVES ON MORAL RESPONSIBILITY 321–347 (J. Fischer & M. Ravizza eds., 1993), at 346. van Inwagen would argue that the agent is not responsible for A or B because, given that he is not responsible for B, he could not have prevented the disjunction, and agents are not responsible for what they could not have prevented. Heinaman argues that an agent can be responsible for the disjunction A or B when he is responsible for A and when B is false, but agrees with van Inwagen that if he is not responsible for B, and B is *true*, then he is not responsible for A or B. Rowe argues that an agent is not responsible for A or B when B is true and it is made true by a completely independent process outside the agent's control. And Fischer and Ravizza argue that an agent is not responsible for A or B when B does not result from a "responsive sequence." So all of these philosophers seem to be endorsing RD.

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

As I explain just below, I do not think that RD is universally true. However, I do think that something *in the neighborhood of* RD must be true and that this closely related principle helps explain our intuitions about certain cases. Consider, in particular, the following examples discussed by van Inwagen.[25] Imagine that Gunnar shoots and kills Ridley and is responsible for doing so. Is he also responsible for the fact that Gunnar killed Ridley or $2 + 2 = 4$? Or for the fact that Gunnar killed Ridley or grass is green? It seems not. If so, a principle like RD would help explain why, since Gunnar is not responsible for the fact that $2 + 2 = 4$ or for the fact that grass is green.

Note that RD entails that the one agent in Two Buttons–One Stuck is not responsible for D. For he is not responsible for the fact that the other button is not depressed (this is the result of a mechanism over which he has no control). Note, however, that RD *also* entails that the two agents are not responsible for D in Two Buttons. For neither agent is responsible for the status of the other button. This is where I think the principle goes wrong. Claiming that the agents are not responsible for D in Two Buttons would violate The Slogan (note that the special circumstances described in Section IV do not apply to the case at hand). Clearly, the agents in Two Buttons *are* responsible for the fact that the buttons are not depressed at the relevant time. Hence, despite what philosophers seem to think, RD is not universally true. It is too strong.

However, I do think that some principle in the neighborhood of RD (a weaker principle) must be true. As van Inwagen's examples seem to suggest, agents are *typically* not responsible for disjunctive facts when they are not responsible for one of the obtaining disjuncts. As far as I can tell, agents are not responsible in circumstances of this kind *unless* absolving them of responsibility would result in a violation of The Slogan. Hence, at least as a first pass, we could try the following revision of the principle:

> (RD*) If both A and B obtain, and an agent is responsible for A, and *no one (neither A nor anyone else) is responsible for B,* then the agent is not responsible for the disjunctive fact: A or B.

Just like RD, RD* only offers sufficient conditions for the lack of responsibility. But unlike RD, RD* is consistent with the truth of The Slogan. As seen above, in Two Buttons RD entails that neither agent is responsible for D, where The Slogan entails that the agents are responsible. In contrast, RD* is silent about the agents' responsibility for D. On the other hand, RD* explains our intuitions about van Inwagen's examples and about Two Buttons–One Stuck (where claiming that the agent is not responsible clearly does not violate The Slogan, since there is only one wrong in those cases, not two). In sum, RD* is a plausible and theoretically fruitful principle of responsibility about disjunctive facts that accounts for our intuitions about

25. van Inwagen, *supra* note 24, at 213–214.

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

the responsibility of agents (or lack thereof) in some paradigmatic cases while remaining consistent with the truth of The Slogan.

I do not think that this is the end of the story, unfortunately. There are further questions that arise for scenarios in which the agents' contributions are asymmetrical in certain ways.[26] The existence of these scenarios suggests that RD* might be too weak. For example, consider a variant on the Two Buttons–One Stuck scenario where the mechanism controlling one of the buttons is a hundred years old. Imagine that the person who installed the mechanism at the time failed to take some necessary precautions that would have prevented the failure of the mechanism today. In that case we would probably want to hold that person responsible for the failure of the mechanism today.[27] As a result, in this case RD* does not entail that the agent in charge of the other button (the agent who fails to depress his button today) is off the hook (the principle is consistent with his being responsible for the disjunctive fact that results in the explosion). I do not think it is obvious what one should say about this case, but the fact that the mechanism is bound to fail, given what happened a hundred years ago, suggests that it might be unreasonable to blame the agent for the explosion today. If so, we should look for a principle that is stronger than RD* (without being as strong as RD) to account for this fact.

I am not trying to offer such a principle here, since it seems clear that potential complications of this kind arise only in connection with scenarios involving asymmetrical contributions by agents. I am restricting my focus to the much "cleaner" symmetrical overdetermination cases, so these complications need not concern us. For our purposes here we may simply stick with D* while restricting the principle's intended application to such cases.[28]

On the basis of both The Slogan and RD*, it is possible to substantiate the claim that the agents in Two Buttons are responsible for the explosion but the agent in Two Buttons–One Stuck is not. The agents in Two Buttons are responsible for the explosion by virtue of being responsible for D, but the agent in Two Buttons–One Stuck is not responsible for the explosion since he is not equally responsible for D. A more general conclusion about the metaphysical bases for responsibility seems warranted: we should conclude that causation and counterfactual dependence are not the relevant responsibility bases in the case of responsibility for outcomes (or at least, they are not the only ones). Being responsible for a cause (or being responsible for something on which the outcome depends) can function as

26. Thanks to Randy Clarke for discussion of this point.

27. Even if that person is already dead. Presumably, we hold Hitler responsible for some things that happen nowadays.

28. Scenarios of asymmetrical contributions are famously tricky (see n. 8 *supra*). Some of the questions that arise for them are: Does the mere fact that one agent acts before the other suggest that their contributions are asymmetrical? If they are asymmetrical, in what way, exactly, are they asymmetrical? (Is the person who acts first more responsible, in virtue of having acted first? Or is the person who acts second more responsible, in virtue of having acted closer to the occurrence of the harm?) It is not at all clear how one should adjudicate these questions.

CAMBRIDGE JOURNALS
http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197

a responsibility basis, and it is the one that is presumably operative in the relevant overdetermination scenarios.

WORKS CITED

Fischer, J., and M. Ravizza, eds. (1993) *Perspectives on Moral Responsibility* (Cornell University Press).

———. (1993) "Responsibility for Consequences." In J. Fischer and M. Ravizza, eds., *Perspectives on Moral Responsibility*, 321–347 (Cornell University Press).

Hall, N. (2004) "The Intrinsic Character of Causation." In D. Zimmerman, ed., *Oxford Studies in Metaphysics*, vol. 1, 255–300 (Oxford University Press).

Heinaman, R. (1986) "Incompatibilism without Alternative Possibilities." *Australasian Journal of Philosophy* 64: 266–276; reprinted in J. Fischer and M. Ravizza, eds., *Perspectives on Moral Responsibility*, 296–309 (Cornell University Press, 1993).

Lewis, D. (1986) "Causation." In *Philosophical Papers*, vol. II, 159–213 (Oxford University Press).

Moore, M. (2009) *Causation and Responsibility* (Oxford University Press).

Rowe, W. (1989) "Causing and Being Responsible for What Is Inevitable." *American Philosophical Quarterly* 26: 153–159; reprinted in J. Fischer and M. Ravizza, eds., *Perspectives on Moral Responsibility*, 310–321 (Cornell University Press, 1993).

Sartorio, C. (2004) "How to Be Responsible for Something without Causing It." *Philosophical Perspectives* 18: 315–336.

Thomson, J.J. (1989) "Morality and Bad Luck." *Metaphilosophy* 20: 203–221; reprinted in D. Statman, ed., *Moral Luck*, 195–215 (State University of New York Press, 1993).

van Inwagen, P. (1978) "Ability and Responsibility." *Philosophical Review* 83: 201–224.

Yablo, S. (2002) "De Facto Dependence." *Journal of Philosophy* 99: 130–148.

Zimmerman, M. (2002) "Taking Luck Seriously." *Journal of Philosophy* 99: 553–576.

CAMBRIDGE JOURNALS

http://journals.cambridge.org Downloaded: 15 Nov 2012 IP address: 68.106.20.197