

Resultant Luck and the Thirsty Traveler*

1. Introduction

There is moral luck to the extent that the moral assessment of agents—notably, the assessment concerning their moral responsibility—can depend on factors beyond their control. Moral luck is considered to be problematic (even if unavoidable, according to some) because it conflicts with what is apparently a very intuitive principle, the principle that we can only be morally responsible for what’s under our control.

Nagel (1979) famously distinguished four varieties of moral luck: causal luck (luck with respect to the causes of our acts), circumstantial luck (luck with respect to the circumstances in which we act), constitutive luck (luck with respect to our own constitution), and resultant luck (luck with respect to the results of our acts). Of the four, resultant luck is typically regarded as *the* most problematic (Nelkin (2013)). Thus it is common to suggest that agents can be responsible for their intentions, or for their attempts, but not for the results of those intentions and attempts, since whether agents can carry out their intentions, or whether their attempts are successful, depends on whether the world cooperates in certain ways, which is beyond the agents’ control (see, e.g., Richards (1986), Thomson (1993), and Wolf (2001)). This view is sometimes inspired by the Kantian conception of the “good will” as the only thing that has value. It rejects the possibility of resultant luck

* Thanks to Juan Comesaña, Brad Weslake, an audience at the University of Arizona, and the participants at a symposium on causation in the law at the Pacific APA. Thanks also to Guglielmo Feis for inviting me to contribute to this issue of *Method*.

while remaining open to the possibility of other forms of luck (or perhaps, even, while *requiring* the possibility of other forms of luck, since the assumption that agents can be responsible for their intentions or attempts may require a commitment to the possibility of the other forms of luck).

In this paper I argue that there is a different and as of yet unexplored reason to think that resultant luck is more problematic than other forms of moral luck. For there is a manifestation of the phenomenon of resultant luck that seems especially puzzling, much more puzzling than the standardly recognized forms of luck. We are told that, to the extent that there is resultant luck, it's because agents cannot control what happens outside the sphere of their will as a result of their acts of will: they do their part by forming certain intentions, or by making certain attempts, and then it's "up to the world" (so to speak) to determine what happens as a result. So the responsibility of agents for results in the world depends on factors beyond their control because what they end up contributing to the world is to a large extent beyond their control. But I'll argue that there is a manifestation of resultant luck that has its source, not at all in the agents' lack of control over their *own* contributions, but exclusively in their lack of control over the contributions of *other* agents or mechanisms. This is a novel and particularly problematic form of resultant luck.

I hasten to add that I don't know if we should take this to show that there is no resultant luck, so I won't be drawing this further conclusion. But I do think that it at least shows that the way in which resultant luck can be problematic (and more problematic than other forms of luck) has not been sufficiently appreciated. And so

a thorough evaluation of the moral luck phenomenon would at least need to take this into account.

I'll illustrate these points with one main example, the example of the thirsty traveler, which has been discussed at great length in the literature on causation in the law. In the next section I introduce the example, and the interesting puzzle to which it gave rise. I'll then propose my own solution to the puzzle, and I'll argue that an important reason the puzzle has proved particularly hard to solve is that the thirsty traveler scenario seems to illustrate the novel form of resultant luck.

2. The thirsty traveler puzzle

There are different versions of the thirsty traveler story.¹ In one of its versions, it goes like this. A man is about to take a trip into the desert. In preparation, he fills his canteen with water. The man has two enemies, X and Y, who want him to die. At T1 X drains the water out of the canteen and refills it with salt (so that it won't feel empty and the traveler won't notice the change). Unaware of what X has done, Y then steals the canteen, at T2, thinking that it's still filled with water. The thirsty traveler dies of thirst at some later time, T3. Two (apparently related) questions arise. First, *the causal question*: Who caused the thirsty traveler's death? Second, *the moral responsibility question*: who is morally responsible for his death? (Of course, in the literature on causation in the law the main question is one about legal responsibility. I won't be concerned with that question here, since my focus is on moral responsibility. From now on, I'll use "responsibility" to mean moral

¹ McLaughlin (1925-6) introduced the original example. It was then discussed in detail by Hart and Honore (1985) and by many others.

responsibility.) Note that the question about moral responsibility presupposes that it is possible to hold agents responsible for outcomes, not just for their intentions or attempts.

It's notoriously hard to give a satisfying answer to those questions. On the one hand, the thirsty traveler wouldn't have died had it not been for the malicious intervention of his two enemies. So it seems appropriate to want to hold them, both of them or at least one of them, responsible for the traveler's death. On the other hand, however, it is unclear how we could hold either of them responsible. For, first, it is hard to see how stealing a canteen filled with salt can causally contribute to the death of someone by thirst. And the same goes for draining the water out of a canteen that will be miles away from the man at the time when he'll need it. And, if neither of the two enemies causally contributed to the death of the thirsty traveler in any way, then how could they be responsible for it?

Note that an implicit assumption behind the way in which I have presented the puzzle is that absence causation is at least possible. In particular, I am assuming that an absence like the lack of water can, in principle, be the cause of a death, so that there is a substantial question about who can bear causal responsibility in this case. This may seem like a reasonable assumption, since it seems to accord with commonsense. However, some philosophers would disagree with this assumption (see, e.g., Beebe (2004)). If, as those philosophers believe, absence causation isn't possible, then clearly the thirsty traveler's enemies could not have caused his death (because, to the extent that they caused it, it must be because they caused something like the absence of water or of the canteen, which in turn caused the death).

However, note that there would still be a puzzle, even in that case, for we think that agents can sometimes be responsible for things in ways that involve absences (they can be responsible for their omissions, and for other types of absences, and they can be responsible for other things in virtue of their responsibility for certain absences, such as their omissions). Thus if absence causation weren't possible, there would still be a puzzle, although the puzzle wouldn't involve causation but some alternative ground for responsibility that applies in cases of that kind. For example, Dowe (2001) introduces the concept of "quasi-causation," which he argues provides an alternative basis for responsibility in scenarios involving omissions and absences in general. Quasi-causation is, roughly, *possible* causation: the absence of water can quasi-cause a death (although it cannot cause a death) to the extent that the presence of water *would have caused* the relevant organism to stay alive. The thirsty traveler puzzle can easily be reformulated in terms of quasi-causation. In what follows, for simplicity's sake, I'll assume that absence causation is possible, and I'll discuss the puzzle in terms of causation. But nothing essential hangs on this.

In the literature there is a wide array of answers to the thirsty traveler puzzle. In fact, for each of the obvious possibilities there is someone who has defended a version of it. The possibilities that most obviously come to mind are these. First, one could argue that X caused the death and Y didn't; as a result, only X is responsible. Second, one could argue that only Y caused the death; as a result, only Y is responsible. Third, one could argue that both caused it and thus both are responsible. Fourth, one could argue that neither caused it and so neither is responsible. Fifth, one could argue that neither caused it but they are still

responsible (or at least one of them is responsible) because responsibility can sometimes fail to be grounded in causation (or in similar relations).²

My proposed solution to the puzzle is different from all of these. It has two parts. The first part is a direct answer to the causal question (what that answer is will be apparent in the next section). The second part is the claim that settling the causal question *fails to settle* the question about responsibility. However, I will argue that this is not because, as the last answer we have reviewed suggests, responsibility in these cases is not grounded in causation. Instead, it is because, although there is an important sense in which responsibility *is* grounded in causation even in these cases, settling the issue about causation is not *sufficient* (in these cases) for settling the issue about responsibility.

Moreover, on the view I'll propose, this is not because of the widely recognized fact that causing a harm is not sufficient to be responsible for it. Philosophers typically identify a set of *epistemic* conditions that are required for responsibility, which include the capacity to foresee that certain consequences would follow from certain acts, the possession of certain beliefs or intentions, etc. Arguably, the agents in the thirsty traveler case satisfy all of these conditions (they

² Gavison, Margalit, and Ullmann-Margalit (1980) have defended the first answer. Mackie (1980) and Hausman (1998) have embraced the second answer, in a variant of the case where a poison takes the place of the salt. Kvat (2002) has argued for the third answer, on a collective interpretation of 'caused' (X and Y together caused the death, but not individually). Moore (2009) has defended the fourth answer. Hart and Honore (1985) seem to (rather reluctantly and hesitantly) endorse the fifth answer, by speculating that this may be a rare case of responsibility without causation. Stapleton (2008) defends a sixth type of solution, according to which there is no single answer to the causal question (there are many different concepts of causation that apply in different contexts) and so there is no hope in trying to settle the responsibility question by settling the causal question; the responsibility question has to be settled in some other way. But Stapleton proposes this as an answer to the question about legal, not moral, responsibility. Hart and Honore's view (the fourth answer described above) may also be best seen as an answer to the question about legal but not moral responsibility.

could foresee that their acts were likely to result in the thirsty traveler's death, they acted with the relevant beliefs and intentions, etc.). So, on the view I'll propose, settling the causal question fails to settle the responsibility question even if all the standard epistemic conditions for responsibility are met in this case. As a result, answering the causal question constitutes *no significant progress* towards answering the responsibility question: the puzzle survives, and as puzzling as ever, as a puzzle about responsibility.³

Thus, I will suggest that a main reason the puzzle has proved so hard to solve is that we have been largely confused about its nature. The thirsty traveler puzzle is not really a riddle about causation, as most philosophers have taken it to be. At least, it's not *just* a riddle about causation, but, more fundamentally, it's a riddle about responsibility. A symptom of this, for me at least, is that I think I know the causal facts now, but I am still quite unsure about the responsibility facts. Although I will make some speculative suggestions about how to go about solving the puzzle about responsibility, I won't attempt to reach any definite conclusions on this matter. My main aim is to argue that what the thirsty traveler scenario reveals is that we have been wrong about the relation between responsibility and causation, to a significant extent. Relatedly, the thirsty traveler scenario reveals that our picture of the phenomenon of resultant luck is in some important ways incomplete.

³ As will be apparent later, my solution is also different from the other answers mentioned in note 2. In particular, it is different from the fourth answer in that I do think that someone *is* responsible in this case and so answering the responsibility question is an urgent matter. And it is different from Kvat's answer (which I catalogued as an example of the third answer) because Kvat is prepared to take the step from collective causation (X and Y are collective causes, although not individual causes) to individual responsibility (X and Y are each individually responsible as a result). As I argue later in the text, I believe that this step is not warranted without further argument.

2. Who killed the thirsty traveler?

Our puzzle began with the causal question: Who caused the thirsty traveler's death?

I will argue that the best answer to this question is: *no one* caused the thirsty traveler's death, although *something* (or the absence of something) did. I will argue for these claims in turn.

When I say that "no one" caused the thirsty traveler's death I mean that neither X nor Y caused it (at least not in virtue of what they did that day). A strong argument for this claim, I think, starts by considering variations of the thirsty traveler scenario where our causal judgments are much clearer, and then extrapolates those same judgments to our case. This is a strategy of argumentation that has rarely been employed in the literature on the thirsty traveler puzzle.

Although some have argued for similar claims, they have argued for it in what I think are less effective ways.⁴ The strategy I'll deploy complements and reinforces those other lines of argument.

The idea is this. Our intuitive causal judgments about the thirsty traveler case itself are quite unclear. So consider, instead, "structurally identical" variations of the thirsty traveler scenario that result from varying the epistemic state of one of the agents involved (his beliefs, intentions, etc.). These variations are "structurally identical" in that, if the only difference (at least the only difference concerning the agent himself) has to do with the agent's epistemic state, then his causal

⁴ An exception seems to be an isolated passage by Hart and Honore, where they write (about a version of the case involving poison instead of salt): "... had [Y] been a well-intentioned doctor intervening to stop [the victim] dying of poison, the natural comment would be that he had not deprived [the victim] of any of the essentials of life, and had incidentally prolonged it slightly" (Hart and Honore (1985), p. 240).

contribution should be the same as in the original case.⁵ So imagine, now, that we can find structurally identical variations where we clearly judge that the agent is/is not a cause, but we cannot find structurally identical variations where we clearly judge that the opposite is true. If so, it seems that we have good reason to believe that the agent is/is not a cause in *all* of these cases, including the original scenario.

This strategy can be applied to the thirsty traveler scenario to show that neither agent is a cause of the man's death. First, to show that X isn't a cause, consider the following variation:

Variation 1: Whereas Y is the traveler's archenemy and wants him to die, X is, in contrast, the traveler's best friend. X is aware of Y's plan to steal the canteen from his beloved friend, who is about to go on a trip in the desert. X has tried to persuade his friend not to take the trip but the traveler is completely determined to do it. The traveler also refuses to believe that anyone might want to kill him, so he fails to take any additional precautions (like safely securing his canteen or carrying extra water). Sadly realizing that he won't be able to prevent his friend's death, and seeking anticipated revenge on the death of his friend, X drains the water out of the canteen and fills it with salt so that at least Y will *also* die of thirst (he knows that Y will be counting on the stolen canteen to survive). As predicted, Y steals the canteen, which is at that point filled with salt, the traveler dies of thirst, and so does Y.

⁵ Compare: I push a button, and an explosion occurs an instant later. Did I cause the explosion by pushing the button? The answer to this question doesn't at all hinge on what I believed, intended, or expected to be the case, but only on the relevant facts of the world.

My causal judgments are very clear in this case: X causes Y's death, but X doesn't cause his friend (the thirsty traveler)'s death. Presumably, in this case X would regard the substitution of salt for water as a way of causing only Y's death, not the traveler's death. And, presumably, this is also how we would regard it from a third-person perspective.^{6 7}

On the other hand, I can think of no variations of the thirsty traveler case that elicit a clear judgment that X *is* a cause. The closest we can come to that are scenarios similar to the original scenario itself, where X has bad intentions (he wants the man to die and substitutes the salt for the water in order to cause his death). But in all those cases the judgment that X is indeed a cause of the man's death is far from clear. So, whereas there are some structurally identical variations of the thirsty traveler case where X is clearly not a cause of the man's death, there are no such variations where X is clearly a cause of the man's death. This suggests that we have good reason to believe that X is not a cause, in all of these cases.

⁶ One difference between the original thirsty traveler scenario and Variation 1 is that here X's act is caused by Y's having planned to steal the canteen. But note that this only generates a difference in Y's contribution towards the death (Y clearly causes the death in this case); X's own contribution towards the death arguably remains the same, since what *he* does is exactly the same (he substitutes salt for water in a canteen that will later be stolen, before the man tries to drink from it). I am using Variation 1 only to show that X's act doesn't cause the death in the original thirsty traveler scenario. (The next variation we'll consider, Variation 2, will be used to show that Y's act doesn't cause the death either.)

⁷ Someone could object that our intuitions are morally tainted in this case. Could it be that X does causally contribute to his friend's death, but we tend to miss this because he's only trying to help, and he carries no blame for the death? Although I agree that our causal judgments are morally tainted sometimes, I don't see the motivation to think this in this particular case. We would need *some* good reason to think that our intuitive judgments shouldn't be trusted in this case, and I don't think there is such a reason. (As I explain later, someone might think that there is a good reason, namely: if X isn't a cause, then neither is Y, and then the man's death appears to be uncaused, when it clearly isn't. I address this worry below in the text.)

A similar argument can be given in support of the claim that Y didn't cause the traveler's death. Consider a different variation of the case:

Variation 2: Now Y is the thirsty traveler's best friend and X is his worst enemy. X has already drained the water out of the canteen and replaced it with salt. Y is aware of what X has done but, unfortunately, there is nothing he can do to warn his friend or to prevent his death. Still, imagine that he can take away his canteen somehow, without his friend noticing. Imagine, further, that, anticipating that his friend would take at least a sip from the canteen before realizing that it contained salt, and believing that the taste of salt in his mouth would make his death even worse (if only by a little), Y takes away the canteen.

Again, my intuitive judgments about this case are very clear: Y doesn't cause the man's death (or, in this case, *any* death, for that matter). Plus, again, I can think of no variations that elicit the opposite clear judgment, the judgment that Y *is* a cause of the man's death. This suggests that we also have good reason to believe that Y isn't a cause of the man's death, in all of these cases. I conclude that there is good reason to believe that neither X nor Y caused the man's death.⁸

At this point, someone might raise the following objection. If this argument were sound, then we could conclude that what the enemies did contributed nothing

⁸ The thirsty traveler scenario is, in this respect, importantly different from a standard (firing-squad type) overdetermination case. Note, in particular, that it is not possible to argue in a similar way for the claim that standard overdeterminers aren't causes.

to the man's death, and so it would follow that *nothing* caused the man's death (besides, say, the man's deciding to go for a walk in the desert). But this is extremely implausible: the man's death is not an uncaused event, or an event that lacks a full explanation. So the argument must be unsound.

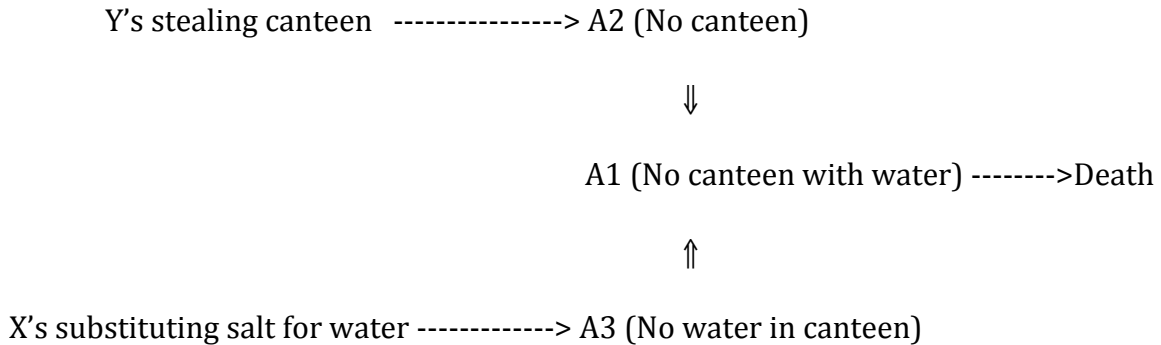
My response is that it *doesn't* follow from the argument above that the man's death is uncaused event, or one that lacks a full explanation. The claim that neither X nor Y caused the death is consistent with the claim that something caused it, and, in particular, with the claim that the cause is something that has to do, in some way or other, with what X and Y did.

So, the question that remains to be answered is: *what* caused the thirsty traveler's death, and how are X and Y connected with that cause? I'll turn to this question now.

Here is one absence that seems to have caused the thirsty traveler's death: the absence of his water-filled canteen from a certain spatiotemporal location, namely, the man's spatial location at time T, the last time when he could have avoided death by drinking water. Note, in particular, that there is counterfactual dependence between this absence and the man's death: if the man had been in possession of his water-filled canteen at that time, then he wouldn't have died. Call this absence 'A1.' Note that the fact that A1 obtains follows from the fact that another absence obtains: the absence of the canteen from the man's spatial location at the same time, T. (If the man was missing his canteen at T, then it follows that he was missing his water-filled canteen at T.) Call this second absence 'A2.' The fact that A1 obtained also follows from the fact that this other absence obtained: the

absence of water from the man's canteen at T, or the canteen's not containing water at T. (If the canteen wasn't filled with water at T, then it follows that the man didn't have in his possession his water-filled canteen at T.) Call this third absence 'A3.'

Finally, note that X's substituting the salt for water at an earlier time caused A3, and Y's later stealing of the canteen caused A2. All of these facts are represented in the following diagram (the dotted arrows are causal relations and the double arrows are logical relations):

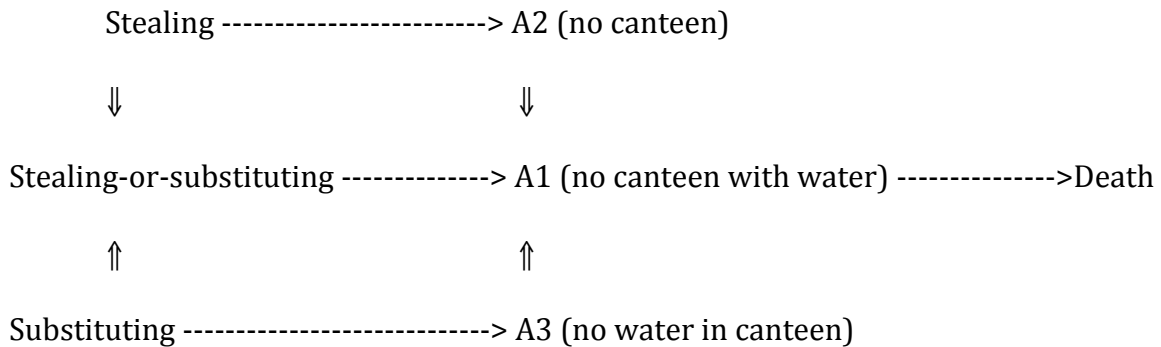


So now we have identified a cause of the death, A1. But, how are the agents' acts related to A1, and thus to the death? We cannot say that they caused A1; otherwise they would have caused the death too (by transitivity, which we don't have a good reason to reject in this case). I think that what we should say is that a disjunctive fact, the "stealing-or-substituting" (the fact that would have obtained just in case either X had substituted the salt for water when he did, or Y had stolen the canteen when he did, or both), caused A1, and thus the death.

This is not the place to defend the possibility of disjunctive causes. I've done this elsewhere.⁹ So, instead of trying to offer a full argument for this here, I will draw attention to just a few main points in support of this idea. First, notice that, once one accepts that absences (such as the lack of water) can be causes, the idea that disjunctive facts can be causes becomes considerably plausible. For, on at least one plausible view of absences, absences are disjunctive facts (for example, my failing to attend the meeting at noon is the fact that obtains just in case I am asleep at the time, or I am at home, or I am at the gym, etc.). Second, note that there is counterfactual dependence between the disjunctive fact, the stealing-or-substituting, and A1: if the stealing-or-substituting hadn't obtained (that is to say, if neither the stealing nor the substituting had taken place), then the thirsty traveler would have been in possession of his water-filled canteen at T (that is, A1 wouldn't have obtained), and thus he wouldn't have died of thirst. At least typically, counterfactual dependence is sufficient for causation, so this is some reason to think that the disjunctive fact caused A1. Finally, although (as we have seen) neither the stealing nor the substituting caused A1, A1 obtained, in some way, *due to* the stealing and the substituting. And there seems to be no other possible way to sustain the claim that A1 obtained due to the stealing and the substituting. (Could one claim, for example, that the event consisting in the mereological sum of the stealing and the substituting caused A1? Arguably not, since it's hard to see how a mereological sum could cause something without any of its parts making a causal contribution.)

⁹ See Sartorio (2006). There I argued for the existence of disjunctive causation in a scenario of a similar kind (one involving two agents acting independently and collectively bringing about an outcome).

Hence I propose that the disjunctive fact, the stealing-or-substituting, caused A1, and thus the death. This completes our diagram of the causal structure of the case:



This might seem very complicated as an account of the causal structure of the case. But it's not, really. The main point that we should extract from it is that the death wasn't caused by the agents' acts themselves; however, it was caused by something *entailed* or *guaranteed* by the agents' acts. Given that the acts themselves didn't cause the death, the agents themselves didn't cause the death (since to say that an agent caused something is arguably just to say that something the agent did or failed to do caused it). However, even if the agents didn't cause the death, they are still connected to the death in an important way, given that, again, something entailed by or guaranteed by their acts caused the death.

This concludes my answer to the causal question. Now let us turn to the question about responsibility.

3. Who is responsible for the thirsty traveler's death?

A short, but quite uninformative, answer to this question is available, now that we have an account of the causal structure of the case. The answer is: whoever is responsible for the cause of the death (the disjunctive fact, the stealing-or-substituting) is also responsible for the death. Note, in particular, that we are assuming that the agents in the thirsty traveler scenario meet all the standard epistemic conditions for responsibility (they acted with a malicious intention, on the belief that their actions would cause the death, etc.). Presumably, then, if they are responsible for the cause of the death, they are also responsible for the death itself.

But, who is responsible for the disjunctive fact? (Is X responsible? Is Y responsible? Are they both responsible?) This question turns out to be very tricky. At first sight, it might be tempting to suggest that we are responsible for anything that logically follows from things for which we are responsible. If that were the case, then both X and Y would be responsible for the disjunctive fact. But, on reflection, there are several problems with this idea. First, the general principle:

If an agent is responsible for P, and Q logically follows from P, then the agent is also responsible for Q.

is clearly false. I am responsible for many of my acts but not for the fact that I exist (or that I ever existed). But the fact that I exist (or that I ever existed) follows from the fact that I perform any given act.

One might think of limiting the scope of the principle, perhaps by restricting it only to disjunctive facts that are entailed by things for which we are responsible.

The more specific principle would say:

If an agent is responsible for P, then the agent is also responsible for P-or-Q, for any Q.

But, first, it's unclear why the more specific principle might be true if the general principle is not. What could possibly make disjunctive facts so special? Second, there are also apparent counterexamples to the more specific principle. For example, arguably, I am not responsible for the fact that either I acted in a certain way (for which I am responsible) or $2+2=4$. Or for the fact that either I acted in that way or grass is green.¹⁰

Finally, and relatedly, the principle also has the consequence that agents are responsible for outcomes in variants of the thirsty traveler scenario where they don't seem responsible. Consider a variant involving only one malevolent agent and some purely natural forces. Imagine, for example, that some fluky, natural phenomenon resulted in the fact that the water in the canteen was replaced with salt. Imagine that, ignoring what happened, Y then stole the canteen. Here Y does *not* seem to be at all responsible for the thirsty traveler's death. In fact, *no one* seems responsible for the unfortunate outcome in this case. If there is resultant luck, we would say that Y is off the hook, and is *lucky* to be off the hook, because the death is

¹⁰ Van Inwagen argues for this in van Inwagen (1978: 213-4).

an unfortunate accident in this case, or an act of nature. But, if the principle that we are considering were true, it would follow that Y is responsible, because he is responsible for the cause of the death.¹¹ (This variant of the thirsty traveler scenario will play an important role in my discussion later. I'll return to it in section 4.)

Note that the original version of the thirsty traveler scenario is importantly different from this new version in that, in the original version, the death wouldn't have taken place had it not been for the malicious intervention of the two agents. The death is, in that sense, an upshot of human wrongdoing. This is an important difference between the two scenarios. And this difference seems to be what's behind the intuition that someone *must* be responsible for the death in the original scenario, even if no one is responsible in the version involving forces of nature.

This rules out one possible answer to the thirsty traveler puzzle: the answer according to which neither X nor Y is responsible. But, of course, it doesn't settle the issue of *who* is responsible (it could be both, or just X, or just Y).¹² The question that remains to be answered is, then: When are we responsible for something that follows from something for which we are responsible? In particular, when are we responsible for a disjunctive fact that follows from something for which we are

¹¹ By the way, this is why I think that Kvat's answer to the thirsty traveler puzzle (see n.2) fails. The agent is as much a "collective cause" of the death in this case as in the original version of the case. But he is not responsible for the death in this case. So we cannot conclude from the fact that an agent is a collective cause of an outcome that he is responsible for the outcome (even if all the standard epistemic conditions for responsibility are met, e.g., the agent had the intention to cause the death, believed that he was causing the death, etc.).

¹² Note that, had the agents acted in completely symmetrical circumstances, then we should have concluded that both are responsible. I discuss a scenario of this sort in Sartorio (2004). Perhaps a variant of the thirsty traveler scenario where the two enemies act simultaneously would be of this kind too (X makes the water drain out of the canteen just as Y is taking it away from the man, or something along these lines). But in the original case it is not clear what we should conclude; that's why there is still a puzzle.

responsible? Here we are walking through basically unexplored territory.¹³ I have a proposal to make, but it is only tentative.

One main thing one would have to do when trying to answer this question is examine possible reasons for thinking that one of the agents might be more responsible than the other, perhaps even to the extent that one of them is exempt from responsibility altogether. I can think of at least one such potential consideration. Although both disjuncts of the disjunctive fact (the stealing-or-substituting) in fact obtain, one of them is of course enough for the disjunctive fact to obtain. Now, X, the agent who substitutes the salt for water acts *before* Y, the canteen thief, and hence guarantees that the disjunctive fact obtains before the canteen thief does. There is a sense in which Y comes too late into the picture: by the time he acts, the disjunctive fact is already guaranteed to occur given what X did. One might think that, in light of this difference, X is more responsible for the disjunctive fact, and thus for the death, than Y—perhaps even to the extent that Y isn't responsible at all.

Since I find this reasoning plausible, I am tentatively drawn to conclude that we should use the temporal priority criterion to answer the question about responsibility in the thirsty traveler case. The answer would then be that X is more responsible than Y, perhaps even to the extent that Y isn't responsible at all, because X guarantees that the cause of the death will obtain before Y does. If the death

¹³ Some philosophers have offered more or less explicit proposals about when agents are/are not responsible for disjunctive facts (see van Inwagen (1978), Heinaman (1986), Rowe (1989), and Fischer and Ravizza (1993)). But I think that all those proposals fail basically because they fail to take into consideration scenarios like the thirsty traveler case (for a discussion of this, see Sartorio (2012)).

happened as a result of that cause, and X's act guaranteed that the cause would occur before Y's act did, then it seems reasonable to conclude that X has a greater claim to be responsible than Y; perhaps, even, that X bears all the responsibility and Y none.

If that's indeed the case, it means that the thirsty traveler scenario illustrates a novel, non-causal form of "preemption": a form of preemption involving responsibility but not causation. There is *causal* preemption when X caused an outcome, Y didn't, but Y would have caused the outcome had X not caused it (see Lewis (1986)). Similarly, let's say that there is *responsibility-involving* preemption when X was responsible for an outcome, Y wasn't, but Y would have been responsible for the outcome had X not been responsible for it. Finally, let's say that there is *non-causal* responsibility-involving preemption (or *pure* responsibility-involving preemption) when there is responsibility-involving preemption without causal preemption. If the criterion of temporal priority establishes that X is responsible for the thirsty traveler's death and Y isn't, then X's responsibility preempts Y's responsibility in this last sense.

Note that it is much more plausible to suggest that temporal priority can give rise to non-causal or pure responsibility-involving preemption than it would be to suggest that it can give rise to causal preemption. The fact that a potential cause occurred before another potential cause is *not* by itself a good reason to think that the former preempted the latter. For example, a slow bullet can be fired at a target before a faster bullet but the faster bullet can still preempt the slower one if the faster bullet reaches the target first. This is so even if the firing of the slow bullet

already guaranteed the destruction of the target (meaning that it was sufficient for it, in the circumstances). For, even if an event guarantees the occurrence of an outcome in this sense, there is still a further fact that determines whether it is indeed a cause of the outcome (the actual trajectory of the bullet, or something of this kind). But, if the question is who is responsible for the cause of a harm, when each of the agents' acts logically guaranteed that the cause of the harm would happen, there is no such further fact to be taken into consideration. The way in which the two agents are causally involved in the situation is, I have argued, the same (what each of them does logically guarantees the occurrence of the cause, and that is the whole extent of their causal involvement vis-à-vis the death). Hence it is much more plausible to suggest that whoever guaranteed the occurrence of the cause *first* bears all the responsibility.

Again, I am only suggesting this as a tentative answer. The main lesson I want to draw from all this is that, in these scenarios, the question about responsibility, far from being settled by the question about causation, remains an *open question* even after we have answered the causal question. In other words, the concept of responsibility is much more "plastic" than the concept of causation, and this is why the puzzle about responsibility survives the resolution of the causal puzzle in the thirsty traveler case.

Notice that I am *not* claiming that the puzzle about responsibility remains because responsibility in these cases is not grounded in causation. It *is* still grounded in causation, in an important way. In these cases, in order to be responsible for the death, you have to be responsible for the relevant cause of the

death (even if you don't have to be a cause of the death yourself). Consider a third enemy, Z, who contributed nothing to the absence of the water-filled canteen. Z is clearly not responsible for the death, and for precisely that reason. Imagine, for example, that Z poisoned the thirsty traveler's food, but the traveler died of thirst before even trying the food. Z clearly bears no responsibility for the death, since he is not responsible for what caused his death. The reason the puzzle about responsibility survives the resolution of the causal puzzle is not, then, that responsibility fails to be grounded in causation in that case. It is, instead, that in these cases settling the causal questions is not *sufficient* for answering the responsibility question.¹⁴ The essence of the puzzle remains because some hard questions persist even after we have answered all the relevant causal questions.

In the literature on causation in the law, some have expressed skepticism about the idea that we can look at what philosophers have said about causation to solve some legal puzzles. Philosophers like Fumerton and Kress (2001) and Thomson (2008) have argued that, whereas the law can draw on intuitions and useful generalizations about causation, no philosophical *theory* of causation presently enjoys (or perhaps will ever enjoy) enough support that the law can draw on it. The problem, they claimed, is not just that there is no widely accepted theory of causation among philosophers (presumably, this is true of any issue that's of philosophical interest!) but, more importantly, the scenarios that the law was hoping to get help with (notably, hard cases such as the thirsty traveler case)

¹⁴ And, as I mentioned before, this isn't just the claim that other conditions are required for responsibility, besides the causal condition. As I have explained, all the typically recognized epistemic conditions are satisfied in the cases we have been considering.

happen to be the ones for which there is no consensus about the causal structure among metaphysicians themselves. If the cases where the law was hoping to get help from metaphysicians are precisely those cases where metaphysicians have trouble being in agreement, then we cannot hope to settle the difficult questions about responsibility by appeal to what metaphysicians have said about causation.

The discussion of the thirsty traveler puzzle helps to bring out that there is another, more fundamental reason to be skeptical about the prospects of looking into the metaphysics of causation to try and settle some questions about responsibility—in particular, questions about tricky cases such as the thirsty traveler case. It is that the right metaphysical view of the matter in these cases may *fail to settle* the questions that law theorists were hoping it would settle. If so, other criteria will have to be brought to bear on these issues, besides causal criteria. And this is, again, not because causation is irrelevant to responsibility or because responsibility fails to be grounded in causation, but because the way in which responsibility is grounded in causation is not the way people have expected it to be. The right way of thinking about the relation between causation and responsibility leaves some important questions about responsibility unanswered, and, in order to answer those questions, we have to “get our hands dirty,” or get them *even dirtier*, and look more closely into the nature of responsibility itself.

4. A new form of resultant luck

Now let me finally return to the issue of resultant luck. The thirsty traveler puzzle is obviously a puzzle about responsibility for *outcomes*. In this context, no one denies

that X and Y are responsible for things like intending to cause the traveler's death or trying to cause his death. The question is who is responsible for the outcome, the traveler's death. That's why so many people naturally thought that the question about responsibility is inextricably tied to the question about causation, "Who caused the traveler's death?"

Now, I have argued that answering the causal question is not sufficient to answer the responsibility question. If this is right, it also helps bring out that there is a new form of resultant luck that has been overlooked in the literature: one that arises in a quite different way from the standardly recognized form, and one that looks particularly problematic, or problematic in some new ways.

In section 3 I introduced a variation of the thirsty traveler scenario where one of the agents is replaced by some purely natural phenomenon. It went like this. First, at T1, some fluky, purely natural phenomenon resulted in the fact that the water in the canteen was replaced with salt. Then, at T2, Y stole the canteen. The thirsty traveler died of thirst at T3. (I'll refer to this kind of variation as a "natural" variation.) I noted that in this case we are tempted to think that no one is responsible for the traveler's death, since we tend to regard it as an unfortunate accident. (Of course, even in this case Y would be responsible for things like intending to cause the traveler's death, or for trying to cause his death, but the point is that he doesn't seem responsible for the death itself.) I also noted that the original scenario is importantly different from the natural version because in the original scenario the death is an upshot of human wrongdoing (it wouldn't have occurred

had it not been for the intervention of the two agents), so in that case we feel that someone *must* indeed be responsible.

Similarly, we can imagine another natural variation of the case where Y, not X, is replaced by a natural phenomenon. Imagine, for example, that after X drains out the water and replaces it with salt at T1, lightning strikes the canteen and vaporizes it at T2. Here, too, we are much more likely to think that no one is responsible for the death. Again, although X seems responsible for intending or trying to cause his death, he doesn't seem responsible for the death itself.

But notice what happens when we compare the agent's responsibility in the original scenario and the natural variants. Someone is responsible for the outcome in the original scenario, but no one is in the natural variants. Imagine, for example, that X is responsible for the death in the original scenario (even if Y isn't, assuming that the temporal criterion discussed in section 3 is correct). But X is not similarly responsible for the death in the natural variant I introduced last, where Y is replaced by lightning. When a natural phenomenon intervenes, X is off the hook; when another human being intervenes, X is responsible. But, of course, whether Y or lightning intervenes is something beyond X's control. It follows that there is a difference in X's responsibility between the two cases, and that difference is completely due to factors that are beyond X's control. In other words, X is subject to resultant luck.

But, it is natural to want to protest, how could a fact of *that* kind make a difference to X's responsibility? Moral luck is puzzling because it is hard to get our minds around the idea that we can be responsible for things due to factors beyond

our control. But, assuming that we manage to get past that point of initial resistance, we can at least understand *how* it is that resultant luck would arise, in the more ordinary cases. We know that, to the extent that we can be legitimately responsible for outcomes in the world, our responsibility is partly determined by our causal contribution to those outcomes. And we know that our causal contribution is largely determined by factors we cannot control—in other words, “causal luck” is widespread. So we can at least understand why it is that, to the extent that we can be legitimately responsible for outcomes in the world, our responsibility will hinge on factors beyond our control. All of this we can understand, because we can understand that responsibility for outcomes is importantly tied to causation.

But now it looks like there are other forms of resultant luck that cannot be accounted for in the same way. They are scenarios where our responsibility for outcomes hinges on factors beyond our control, but where those factors *don't* concern our own causal contribution to those outcomes, but the contribution of *other* agents or natural mechanisms. And this seems very odd. While we understand how the connection between causation and responsibility can be a potential source of resultant luck, this other source of resultant luck is different; it cannot be explained in the same way.

In fact, as we have seen, what gives rise to this new form of luck is an unexpected *break* between causation and responsibility. There can be this type of luck because the issue of responsibility can be up for grabs even after we've settled the issue of causation; in other words, because settling the causal facts may not be enough to settle the responsibility facts. The thirsty traveler scenario suggests that

this can in fact happen. And then all that's needed for the new type of resultant luck to arise is the contrast between the original scenario and the natural variations. Although the agent's own causal contribution is the same in both cases, we see that his responsibility can vary depending on whether he's in the original scenario (where he seems responsible) or in a natural variation (where he doesn't seem responsible). That's how we end up with a difference in responsibility without a difference in causal contribution. This interesting phenomenon is exclusive to resultant luck. It suggests that responsibility for outcomes is a more complex phenomenon than it appears to be at first sight.¹⁵

I conclude that a commitment to responsibility for outcomes seems to carry with it a commitment to the possibility of a novel and particularly puzzling form of resultant luck. This commitment would, of course, be avoided by a view that restricted our responsibility to pure intentions or attempts. Even if I'm not sure that this is the conclusion that we *should* draw from all this, it is certainly an important consideration in evaluating the intelligibility of a concept of responsibility that applies to outcomes in the world, and not just to the formation of intentions or the making of attempts.

¹⁵ In Sartorio (forthcoming) I discuss an apparent manifestation of a similar phenomenon for *causal* luck. But I argue that some of the intuitions that give rise to it can be rejected quite easily (in contrast, I don't think that there is any easy way to reject them in scenarios like the thirsty traveler). I also discuss another apparent manifestation of the phenomenon for resultant luck, one that concerns degrees of responsibility.

References

- Beebe, H. (2004) "Causing and Nothingness." In J. Collins, N. Hall, and L. A. Paul (eds.), *Causation and Counterfactuals* (Cambridge, Mass: M.I.T. Press), pp. 291-308.
- Dowe, P. (2001) "A Counterfactual Theory of Prevention and 'Causation' by Omission." *Australasian Journal of Philosophy* 79, 2: 216-26.
- Fischer, J. and M. Ravizza (1993) "Responsibility for Consequences." In J. Fischer M. and Ravizza (eds.), *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press), pp. 321-47.
- Fumerton, R. and K. Kress (2001) "Causation and the Law: Preemption, Lawful Sufficiency, and Causal Sufficiency." In *Law and Contemporary Problems* 64, 4: 83-105.
- Gavison, R., A. Margalit, and E. Ullmann-Margalit (1980) "Causal Over-determination: Resolution of a Puzzle." *Philosophical Studies* 37 (4): 381-90.
- Hart, H. L. A. and T. Honore (1985) *Causation in the Law*, second ed. Oxford: Oxford University Press.
- Hausman, D. (1998) *Causal Asymmetries*. Cambridge: Cambridge University Press.
- Heinaman, R. (1986) "Incompatibilism without Alternative Possibilities." *Australasian Journal of Philosophy* 64: 266-76.
- Kvart, I. (2002) "Probabilistic Cause and the Thirsty Traveler." *Journal of Philosophical Logic* 31: 139-179.
- Lewis, D. (1986) "Causation." In *Philosophical Papers*, vol. II (New York: Oxford U.P), pp. 159-213.
- Mackie, J. L. (1980) *The Cement of the Universe*. Oxford: Oxford University Press.

- McLaughlin, J. A. (1925-6) "Proximate Cause." *Harvard Law Review* 39: 149-55.
- Moore, M. (2009) *Causation and Responsibility*. New York: Oxford University Press.
- Nagel, T. (1979) "Moral Luck." In *Mortal Questions* (New York: Cambridge University Press), pp. 24-38.
- Nelkin, D. (2013) "Moral Luck." *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), E. N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2013/entries/moral-luck/>.
- Richards, N. (1986) "Luck and Desert." *Mind* 65: 198-209.
- Rowe, W. (1989) "Causing and Being Responsible for what is Inevitable," *American Philosophical Quarterly* 26: 153-9.
- Sartorio, C. (2004) "How to be Responsible for Something without Causing It," *Philosophical Perspectives* 18, 1: 315-36.
- Sartorio, C. (2006) "Disjunctive Causes." *Journal of Philosophy* 103, 10: 521-38.
- Sartorio, C. (2012) "Two Wrongs Don't Make a Right: Responsibility and Overdetermination." *Legal Theory* 18: 473-90.
- Sartorio, C. (forthcoming) "A New Form of Moral Luck?" In A. Buckareff, C. Moya, and S. Rossell (eds.), *Agency and Moral Responsibility* (Palgrave-Macmillan).
- Stapleton, J. (2008) "Choosing What We Mean by "Causation" in the Law." *Missouri Law Review* 73: 433-80.
- Thomson, J. (1993) "Morality and Bad Luck." In D. Statman (ed.), *Moral Luck* (Albany: State University of New York Press), pp. 195-215.
- Thomson, J. (2008) "Some Reflections on Hart and Honore, *Causation in the Law*." In *The Legacy of H. L. A. Hart* (Oxford: Oxford University Press, 2008), pp. 143-65.

Van Inwagen, P. (1978) "Ability and Responsibility." *The Philosophical Review* 83, 2:
201-24.

Wolf, S. (2001) "The Moral of Moral Luck." *Philosophic Exchange* 31: 4-19.