

Responsibility without Causation,
Luck, and Dying of Thirst: A Reply
to Sartorio

MATTHEW TALBERT

West Virginia University
Matthew.Talbert@wvu.edu

Pages: 173 – 184

Keywords: Thirsty Traveler, Sartorio, Causation, Moral Luck, Moral Responsibility

Abstract

This reply to Carolina Sartorio’s “Resultant Luck and the Thirsty Traveler” begins with a discussion of earlier treatments of the thirsty traveler puzzle. I emphasize the way in which adjustments to the case can elicit varying intuitions and conclude with a suggestion as to why the case is so difficult to analyze. Next, I turn to Sartorio’s analysis of the puzzle. I largely agree with her judgments about the causal issues in the case but I am less certain about the conclusions she draws with respect to moral responsibility. I conclude with some reflections on Sartorio’s claim that causation and moral responsibility come apart in the thirsty traveler puzzle. I suggest that this might be true in a way other than the one that Sartorio has in mind; my proposal raises issues about moral luck in addition to those that Sartorio considers

I am delighted to have the opportunity to comment on Professor Sartorio's fascinating and powerfully argued paper. I'll begin with a discussion of earlier treatments of the thirsty traveler puzzle, emphasizing the way in which adjustments to the case can elicit varying intuitions; next, I'll assess Sartorio's treatment of the case; finally, I'll consider Sartorio's suggestion that causation and moral responsibility can come apart and I'll suggest a different sense in which this might be true.

1 The Thirsty Traveler

Here is James McLaughlin's original version of the case:

Suppose the traveler is entering a desert. X secretly empties the traveler's water keg and weights it with salt. The traveler takes the keg into the desert where Y steals it, both the traveler and Y thinking it contains water. The traveler dies of thirst. Who killed him? (McLaughlin 1925, p. 115 note 25).¹

In *Causation in the Law*, Hart and Honoré present an altered version of the case, claiming that the change they introduce "makes the difficulties [presented in the case] clearer" (Hart & Honoré 1985, p. 239, note 70).² As it happens, I think Hart and Honoré are mistaken about this: the real difficulty in the case is most apparent when we consider a version like McLaughlin's (or Sartorio's, which preserves the essential features of McLaughlin's case). At any rate, the difference Hart and Honoré introduce is that in their version of the case, X "secretly puts a fatal dose of poison in" the traveler's water keg (instead of replacing the water with salt) before Y steals it (Hart & Honoré 1985, p. 239).

With respect to this version of the case, Hart and Honoré argue that we cannot confidently answer McLaughlin's question ("Who killed the traveler?") because Y's action neutralizes X's and vice versa. When Y steals the water keg, he prevents X's plan from coming to fruition since the traveler will no longer die from drinking the poisoned water (Hart & Honoré 1985, p. 240). However, we might also say that X pre-emptively neutralizes Y's plan because, given that X has poisoned the water, "it is not possible to describe [Y]'s later action as causing [the traveler's] death" (Hart & Honoré 1985, p. 240). According to Hart and Honoré, to say

¹This is not an exact quotation as I have imported Sartorio's names for the various players; otherwise, the words are McLaughlin's. Since there is no agreement among the authors I discuss in terms of how they refer to the actors in the puzzle, I will use Sartorio's names throughout this paper.

²I am quoting the second edition of *Causation in the Law*; however, Hart and Honoré present largely the same account of the case in the earlier 1959 edition of the book. The main change in the second edition is a discussion of J. L. Mackie's treatment (see below) of the version of the case described in the first edition.

that Y caused the traveler's death "presupposes that the victim had or would, but for the actor's intervention, have had access to the necessary food or drink" (Hart & Honoré 1985, p. 240). But since X poisoned the water, Y did not deprive the traveler of anything life-sustaining when he stole it. Indeed, in preventing the traveler from drinking the poisoned water, Hart and Honoré assume that Y extended the traveler's life,³ but "causing death' involves the notion of shortening the span of life" that a victim would otherwise have had (Hart & Honoré 1985, 240).

Hart and Honoré appear to assume that X uses a relatively fast-acting poison, perhaps one that acts so quickly (or is so strong) that it prevents the water to which it was added from performing its normal hydrating and life-sustaining functions. But this need not be the case. Suppose that X used a slow-acting poison that does nothing to prevent the traveler from being hydrated in the normal way by the water he consumes, even though the poison in the mixture will eventually give rise to other bodily effects that will certainly kill the traveler (though perhaps much later). In this case, Y's theft of the water neutralizes X's effort to bring about the traveler's death by poisoning, but the reverse is not so clearly true: the presence of the poison does not prevent Y from bringing it about that the traveler dies of thirst on account of lacking his water-filled canteen. It seems to me fairly natural in this version of the case to say that Y killed the traveler since the traveler will die sooner than he would have (and in a different way than he would have) if Y had not acted.

J. L. Mackie takes roughly the approach just described. In Mackie's version of the case, Y punctures the traveler's water can instead of stealing it, and Mackie argues that Y caused the traveler's death since his death was realized by the causal chain "puncturing-lack-of-water-thirst-death" rather than by "the rival chain that starts with poison-in-can" (Mackie 1980, p. 46). This is so even if we cannot say that Y's puncturing the can caused the general fact *that the traveler died on his journey* to obtain, since this fact would have obtained regardless of Y's action (Mackie 1980, p. 46). Rather, Y caused the fact *that the traveler died of thirst on his journey* to obtain, which allows us to say that Y caused the "concrete event" of the traveler's death from lack of water (Mackie 1980, p. 46).⁴

Hart and Honoré reply to Mackie by emphasizing the fact that Y merely deprived the traveler "of something lethal, viz. of poisoned water, which does not sustain life" (Hart & Honoré 1985, p. 241 note 74). Thus, for Hart and Honoré, the traveler's case is different from other over-determination cases. Suppose that X starts a fire that will certainly destroy the traveler's house but that before it does

³They say, for example, that Y "has in fact removed one cause [of death] and secured that another cause (thirst) takes effect *later*" (Hart & Honoré 1985, p. 240 emphasis added).

⁴The point could perhaps also be put in terms of a distinction between general and specific consequences: it might not be right to say that Y caused the consequence that the traveler died *one way or another*, but he did cause the more specific outcome that the traveler died of thirst.

so, the fire is quenched by waters that Y released from a damn with the aim of destroying the traveler's house and that these waters in fact destroy the house (cf. Honoré 1985, p. 239). In a scenario like this, Hart and Honoré have no trouble concluding "that the party whose subsequent act neutralized the first act [that is, Y] has alone caused the harm" (Hart & Honoré 1985, p. 239). But the thirsty traveler case is different, Hart and Honoré maintain, because in stealing the traveler's water, Y "neither initiated a series of physical changes in [the traveler's] body nor deprived him of the essentials of life" (Hart & Honoré 1985, p. 241). Of course, this just brings us back to a point made above; namely, that the proper analysis of the case hangs—at least in part—on whether the contents of the traveler's water can be life-preserving in the way that water normally is. It may be that the water in the can, though mixed with a deadly dose of poison can still do the job of sustaining life (to the degree that, and in the way that, water usually does), and in this sort of case, I suspect that Mackie's analysis is the right one.

Gavison, Margalit, and Ullmann-Margalit emphasize the point I just made. For them, what we should say about the case partly depends on issues like the following:

Was the poison used of such type, quantity, and degree of concentration as to still permit the description of the resultant mixture . . . as *water-cum-poison*? That is, can one who drinks it still be plausibly described as having consumed water, albeit that his systems are being fatally damaged in the process. [. . .] Alternatively, the situation may be such that . . . [the mixture] can by no means be considered as poisoned water. (Gavison et al. 1980, p. 389)

In places, Gavison et al. write as if they reject Mackie's analysis, but in fact they seem to accept it, at least for the specific version of the case that Mackie appears to envision. That is, Gavison et al. are prepared to accept Mackie's conclusion in the case in which Y (and only Y) deprives the traveler of *water* (Gavison et al. 1980, p. 387). However, if the case were such that we were inclined to say that X had *already* deprived the traveler of water (by adding poison of a certain sort), then Gavison et al. conclude that it is X who caused the traveler's death (Gavison et al. 1980, p. 389–390).

I'm not sure about this last conclusion since even in this version of the case, Y may ensure that the traveler dies of thirst rather than by poisoning. But perhaps another version of the case comes closer to definitively presenting X as the cause of the traveler's death. Suppose that X saturates the traveler's water with salt, thus ensuring that he dies of thirst in the desert (Gavison et al. 1980, p. 385).

It's tempting to agree with Gavison et al. that X causes the traveler's death here since X has already ensured the traveler's death (as well as the way he dies) before Y has a chance to act. And given that X ensures precisely the sort of death that

Y attempts to bring about, it seems unlikely that Y's action can be construed as neutralizing X's. However, at least for me, the temptation to think that X caused the traveler's death (in the revised case involving salt instead of poison) is lessened by reflecting on what Sartorio calls "natural" variations of the case: suppose that the traveler's saltwater-filled canteen is vaporized by a lightning strike or perhaps that he simply lost the canteen in the desert (Sartorio 2015, p. 168). In these versions, it's difficult for me to sustain the intuition that X causes the traveler's death. But I don't think that the fact that the canteen was destroyed by lightning (rather than being stolen) should matter as far as X's causal contribution to the traveler's death goes, so for me these natural variations also raise doubts about X's causal contribution in the original saltwater case. (This last point will be important for my analysis of Sartorio's approach to the case.)

One thing I find particularly vexing about the thirsty traveler puzzle is that I am not only unsure exactly how to answer McLaughlin's original question, I'm not even sure I can say why that question is so difficult to answer. However, here is at least something one might say about this last issue. Part of the problem (at least in the saltwater version of the case) seems to be that what is over-determined is an *absence* of water. In some other over-determination cases—such as the one featuring the house that will be destroyed by fire or flood—the actual causal route to an outcome will leave an obvious trace in that outcome that informs our causal judgments: it will often be clear whether it was the fire or the flood because the outcomes brought about by these two means would be different in their particulars (e.g., if the house is destroyed by the flood, there will be a lot more water around) even if these outcomes are the same at a more general level of description (i.e., that the house is destroyed). But an absence of water is just an absence no matter how it is brought about, and this makes it difficult to find something distinctive in the outcome that helps us uncover its source. It may, for example, be difficult to point to something that would be distinctive of the absence in question if one agent had caused it, and that would be different if the other agent had caused it.

2 Sartorio's Analysis

The version of the case that Sartorio treats is similar in relevant respects to the version from Gavison et al. that I introduced at the end of the last section: X empties the water from the traveler's canteen, replaces it with salt, and at some later point, Y steals the canteen and the traveler dies of thirst. Sartorio argues that neither X nor Y caused the traveler's death. It certainly is hard to see how Y could have caused the death since, as Sartorio puts it, it is unclear "how stealing a canteen filled with salt can causally contribute to the death of someone by thirst" (Sartorio 2015, p. 156). And "the same goes," Sartorio says, "for [X's] draining the water

out of a canteen that will be miles away from the man at the time when he'll need it" (Sartorio 2015, p. 156).

Sartorio's first point seems to me unassailable (so I am fairly confident that Y did not cause the traveler's death); the second point is perhaps more open to question. One might argue, for example, that the fact that the traveler's canteen was far away at the time of death would be relevant only if—as in the version of the case featuring poisoned water—the presence of the canteen played a role in X's plan. Presumably, the traveler needs to believe that his canteen is nearby up to a certain point in his journey (the point at which he would die of thirst even if he discovered his lack of water and tried to return to wherever his journey began), but after that point, the location of the canteen is immaterial to whether X achieves his goal. (Indeed, the location of the canteen is immaterial even at an earlier point, so long as the traveler *believes* he has the canteen). On the other hand, and to return to the observation with which I ended the last section, perhaps Y's having removed the canteen from the traveler's vicinity makes us think that the absence of water that kills the traveler is somehow a different absence than the one that would have obtained had the traveler's (useless) canteen been nearby.

At this point, it seems fair to say that our intuitions about who caused the traveler's death are uncertain. Thus, Sartorio suggests that we would do well to reflect on related cases. The cases Sartorio has in mind are meant to be identical to the original case in terms of the relevant features of their causal structures, though they may differ in causally irrelevant ways (e.g., the motives of the agents may vary from the original case) or in ways that are causally significant but irrelevant with respect to the point Sartorio means for the case to illustrate. Sartorio's view is that if our causal judgments are clearer in these new cases than in the original case, then, since the new cases share the original's causal structure (in the relevant respects), we would have grounds for applying these relatively clear judgments to the original case (Sartorio 2015, pp. 160–161). This is a strikingly elegant argumentative strategy.

Sartorio believes that we can formulate cases in which X clearly *does not* cause the traveler's death and other cases in which the same is true of Y, yet we cannot find relevant cases in which it is clear that either X or Y caused the death, so we should conclude that neither X nor Y caused the death in the original case. Since I find X a more plausible candidate for having caused the traveler's death than Y, I'll focus on the case Sartorio offers as a clear example in which X does *not* cause the death. In this version of the case, X is the traveler's best friend and he knows that there is no way to dissuade his friend from traveling or of preventing Y from stealing the canteen once the traveler has reached the appropriate point in his journey (Sartorio 2015, p. 161). As a result, X drains the water from his friend's canteen so as to deprive Y of the water that he plans to steal. X does this in order

to ensure Y's death since he knows that Y will rely on the stolen water for his own survival.

Is it clear that X does *not* cause the traveler's death in this case? If it is, then one worry we might have is that this is because Y clearly *does* cause the death (which appears to be Sartorio's view (Sartorio 2015, p. 162 note 6)). This threatens to be a problem for Sartorio's argumentative strategy insofar as that strategy relies on there being no case in which Y (or X) clearly causes the death. However, the difficulty here is only apparent since in the revised case, Y's causal contribution has been significantly altered from the original case (while X's is kept the same), so we cannot apply the judgment that Y caused the death to the original case.⁵

On the other hand, we may not agree that the revised case is one in which X clearly does not cause the traveler's death. There is, no doubt, a sense in which X does not intend the death, and as Sartorio points out, X may not regard himself as the cause of the death (Sartorio 2015, p. 162), but perhaps X is wrong in this causal judgment. Of course, the reader who takes this route is probably also inclined to say that X caused the traveler's death in the original case, and Sartorio's argument isn't meant for people who are strongly inclined toward this conclusion. After all, one who thought it likely that X caused the traveler's death in the original case could perfectly well take a hard-line approach to the revised case: that is, one might conclude that since X caused the death in the original case, and the revised case is relevantly causally identical to the original, then X also causes the traveler's death in the revised case.⁶

Setting the preceding considerations aside, it seems to me that if we agree with Sartorio that neither X nor Y caused the traveler's death in the original case, then her ingenious account of what *did* cause that outcome is very plausible. The account starts with the observation, which is surely correct, that the traveler's death was caused by "the absence of his water-filled canteen from . . . the man's spatial location at time T, the last time when he could have avoided death by drinking water" (Sartorio 2015, p. 164). Sartorio calls this absence A1. What caused A1? If (as the argument now assumes) neither X nor Y caused the traveler's death, then we can't say that either of them caused A1, for then one of them would have caused the traveler's death by having caused A1.

On the other hand, both X and Y caused *other absences* to obtain: Y caused the absence of the canteen from the traveler's location (A2) and X caused the canteen to lack water (A3) (Sartorio 2015, pp. 160–161). Again, we cannot say that either A2 or A3 caused A1 (for in that case either X or Y would have caused the traveler's

⁵I'd like to thank Sartorio for making this aspect of her argument clear to me in conversation. One might worry, though, that if Y's causal role has changed in the revised case, this change also infects X's causal contribution (since the change to Y's role causes X to act as he does).

⁶I call this a hard-line approach because it is similar to the approach that Michael McKenna advocates in "A Hard-Line Reply to Pereboom's Four-Case Argument."

death), but A2 and A3 are obviously connected to A1. In particular, A1 follows logically both from A2 and from A3: that is, the absence of the water-filled canteen (A1) follows logically both from the absence of the canteen (A2) and from the lack of water in the canteen (A3). Likewise, it logically follows from the actions of both X and Y that the traveler either had a canteen in which salt had been substituted for water or he lacked a canteen altogether. On Sartorio's account, it is the obtaining of this "disjunctive fact" (the fact of "stealing-or-substituting") that causes A1; so, while neither X nor Y caused the traveler's death, it is still the case that "something entailed by or guaranteed by their acts caused the death" (Sartorio 2015, p. 167).

While Sartorio believes that neither X nor Y caused the traveler's death, she finds it plausible that there are grounds for saying that X is morally responsible for the death to a degree that Y is not; indeed, X might be solely morally responsible for the death (Sartorio 2015, p. 164–165). The basis for X's greater (or sole) responsibility is that actions of his guarantee that the disjunctive fact described above obtains before Y has the opportunity to act. If Sartorio is correct in her analysis, the question of moral responsibility in the thirsty traveler case remains open even after the causal question has been answered (Sartorio 2015, p. 167).

Now comes the issue of moral luck, and it is here that my intuitions begin to depart significantly from Sartorio's (though she is herself not fully committed to a specific conclusion in this context).

Recall Sartorio's "natural" version of the puzzle. In this version, a lightning strike vaporizes the canteen at the time when Y would have stolen it. Sartorio believes that in this case readers are likely to say that X is *not* morally responsible for the traveler's death, though he is responsible in the original case (Sartorio 2015, p. 168). This introduces a form of resultant moral luck that is distinct from more well known varieties since X is responsible if Y removes the canteen from play, but not if this removal results from natural phenomena. This form of moral luck depends not just on whether *something* beyond our control interferes with a causal process that we initiated, but with whether this something happens to be a moral agent.

For my own part, I don't see how the interference of the lightning strike with X's plans can be relevantly different from Y's interference (at least insofar as these things have an effect on what we want to say about X). Thus, I feel significant pressure to deny that whether the traveler's canteen is removed by Y or by a lightning strike can change what I say about X's responsibility in this case. However, the natural version does incline me to question X's responsibility, but since I doubt that the natural version of the case is different from the original version (in a way that's relevant to X's responsibility), this makes me question X responsibility in the original case.

I suggested in the last section that it's the natural variation of the case that,

for me, really loosens the grip of the conclusion that X *caused* the traveler's death in the original case; now the same variation also seems to loosen the grip of the conclusion that X is *responsible* in that case. Perhaps what this means is that I do not really agree with Sartorio that we have here a case in which the facts about causation do not fix the facts about responsibility—though this is an idea that I initially found attractive (for reasons that will soon become apparent).

3 Moral Luck and Blameworthiness

As we've seen, Sartorio's approach calls into question (in a certain way) the expected link between responsibility and causation. I am attracted to an account of moral responsibility that also questions this link, though in a different way. On the account I favor, assessments of moral responsibility are ultimately attuned (or at least ought to be) to our perception of how a person is as a moral agent and how she is oriented toward other moral agents. A person's actions can, of course, provide us with vital insight into her moral orientation: thus, an agent is blameworthy *for an action* to the degree that it can be attributed to her as an expression of objectionable evaluative judgments regarding, among other things, the needs and interests of the other agents. Typically, actions will be attributable to an agent in this way only if she satisfies a control condition; that is, insofar as she controls her behavior in light of her judgments.

On the other hand, blame-grounding judgments can obtain independently of an agent's controlled behavior. For example, it may be that an agent's unbidden attitudes, or her uncontrolled advertences and inadvertences, express her evaluative orientation just as well as her actions and thus provide similar grounds for blame. More generally, an agent may not have causally contributed (in a way that could ground responsibility) to the fact that she is now the sort of person who is inclined toward objectionable evaluative judgments—yet we may still think that she is open to the negative moral assessments and moral objections that characterize blame simply on account of being the sort of person that she is. Such an agent is not causally responsible for the faults she has, but the faults in question may still be attributable to her for the purposes of moral appraisal.⁷

So, on this account, an agent can be blamed for what is beyond her control if the possession of certain moral judgments (which are not necessarily the product of control) is sufficient for blameworthiness. If we go further and say that possession of such judgments is *all* that matters for blameworthiness, then this position entails the absence of resultant moral luck: the degree to which an agent's environment cooperates with her efforts to shape the world according to her evaluative

⁷For a loosely associated set of views that have some of the features described here, see, among others, Adams (1985), Scanlon (2008), Smith (2005), Talbert (2013).

judgments does not affect her blameworthiness since her blameworthiness is already fixed by the presence of these judgments.⁸ (Note that this is different from the position that eliminates moral luck by insisting that we are responsible only for what is in our control.)

Unfortunately, the position just sketched conflicts with some fairly stable intuitions about moral responsibility (as I suspect will be true of any position on the subject of moral luck). For one thing, the degree to which we blame others is often clearly calibrated to their actual behaviors and to the causal consequences of those behaviors. It may be, though, that the thirsty traveler puzzle gives us a case in which even someone who is undecided about resultant moral luck will find it intuitive to judge that an agent's causal contribution is unrelated to his moral blameworthiness. I'm not suggesting merely that (as Sartorio agrees) we might hold both X and Y responsible for their bad intentions independently of whether their actions caused the traveler's death. Instead (or in addition), I am suggesting that if we agree that X and Y are blameworthy on account of their intentions and choices, then concluding that either X or Y was actually the cause of the traveler's death would do little to change the basis for blaming either agent or the intensity of the blaming responses to which they are open.

Consider the traveler's surviving relatives and how they might respond to X and Y. Presumably, it's reasonable for the relatives to direct blaming responses at X and Y even if they regard the question of who actually killed the traveler as unsettled. I take it that the same would be true if the relatives believed (on grounds like Sartorio's) that neither X nor Y caused the traveler's death. But now suppose that the relatives—who are oddly interested in philosophical accounts of causation—come to believe that X and Y jointly caused the death, or that X rather than Y did it, or the other way around. If nothing else about the case changed, it's not clear to me that the relatives would have (or would regard themselves as having) learned something that gave them access to additional blaming responses or to more intense versions of the responses that were already legitimate.⁹

Whereas an agent's status as the cause of an outcome usually seems concretely related to his blameworthiness, it seems curiously independent of the viability and justifiability of blaming responses in the present case, presumably because the causal issues here are so uncertain and difficult to analyze. (Likewise, if Sartorio is right in attributing a kind of moral responsibility that floats free of causation in X's case, it is not clear to me that this form of responsibility is associated with increased blameworthiness.) Of course, if either X or Y (but not both) caused the

⁸But note that while Scanlon has a view along the lines just described, he attempts to accommodate resultant moral luck (Scanlon 2008, pp. 148-152).

⁹I am thinking here only of the relatives' personal blaming responses, characterized mainly by the expression of certain reactive emotions—I don't take my point to have any necessary bearing on either criminal or civil legal liability.

traveler's death, then the death is on that actor's moral account in a way that it is not on the other actor's. But, again, I don't think this necessarily tells us anything that should be reflected in our moral responses to the causally responsible agent because it doesn't tell us anything that reflects more poorly on him as a moral agent than would otherwise be the case. Of course, I'm just focusing on the kinds of considerations that one would expect a skeptic about moral luck to emphasize, but it does seem to me that these considerations are especially conspicuous in the thirsty traveler case.

References

- Adams, R. M. (1985), 'Involuntary sins', *Philosophical Review* 94(1), 3–31.
- Gavison, R., Margalit, A. & Ullmann-Margalit, E. (1980), 'Causal overdetermination: Resolution of a puzzle', *Philosophical Studies* 37(4), 381–390.
- Hart, H. L. A. & Honoré, T. (1985), *Causation in the Law*, Oxford University Press, New York.
- Mackie, J. L. (1980), *The Cement of the Universe: A Study of Causation*, Clarendon Press.
- McKenna, M. (2008), 'A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument', *Philosophy and Phenomenological Research* 77(1), 142–159.
- McLaughlin, J. (1925), 'Proximate cause', *Harvard Law Review* (39), 149–199.
- Sartorio, C. (2015), 'Resultant Luck and the Thirsty Traveler', *Methode* 4(6), 153–171.
- Scanlon, T. (2008), *Moral Dimensions: Permissibility, Meaning, Blame*, Belknap Press of Harvard University Press.
- Smith, A. M. (2005), 'Responsibility for attitudes: Activity and passivity in mental life', *Ethics* 115(2), 236–271.
- Talbert, M. (2013), Unwitting wrongdoers and the role of moral disagreement in blame, in D. Shoemaker, ed., 'Oxford Studies in Agency and Responsibility Volume 1', Oxford University Press, pp. 225–245.