

# Preemptive Prevention, Causal Difference, and Its Implications\*†

Kim Sungsu

Some causal situations appear similar but causally different. Preemptive prevention offers such an example. However, it is difficult to capture the causal difference. In fact, it resists analysis. I discuss Collins's account of causal difference, and argue that it is not adequate. I also argue that the causal difference is not analyzable in terms of either dependence or production. I then examine some non-reductive account of causal difference, and claim that the sense of causal difference has yet to be specified. Preemptive prevention shows that the concept of causation, whether it is reducible to something else or not, still requires some further scrutiny.

**Subject** Analytic Philosophy, Metaphysics

**Key words** preemptive prevention, causal difference, concepts of causation, Collins

---

\* Received: Jan. 21, 2016 Revised: Feb. 16, 2016 Accepted: Feb. 18, 2016

† This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government. (NRF- 2012S1A5A8023935).

I thank anonymous referees of this journal for their helpful comments and suggestions

## 1. Introduction

Some causal situations appear similar but causally different. Consider the following cases:

Two Catchers: As the ball flew toward us, I leapt to my left to catch it. But it was you, reacting more rapidly than I, who caught the ball just in front of the point at which my hand was poised. Fortunate for us that you made the catch. The ball was headed on a course that, unimpeded, would have taken it through the glass window of a nearby building. Your catch prevented the window from being broken. (Collins 2000, p. 223)

Wall: Suppose that I reach out and catch a passing cricket ball. The next thing along in the ball's direction of motion was a solid brick wall. Beyond that was a window. Did my action prevent the ball hitting the window? (McDermott 1995, p. 525)

Clearly, your catch prevented the window from being broken in Two Catchers. Without your catch, my catch would have prevented the outcome, but your catch preempted mine. This is a case of preemptive prevention. What about Wall? Did my catch prevent the ball hitting the window? It seems not. Or it is not clear whether my action did so. McDermott reports, "Nearly everyone's initial intuition is, 'No, because it wouldn't have hit the window irrespective of whether you had acted or not'" although he quickly adds that "nearly everyone then retracts his initial intuition" when he asked which one, me or the wall, prevented the ball hitting the window (Ibid). Collins, on the other hand, sides with 'everyone's initial intuition': "Given that the wall was there, the window was never in any danger of being broken. The presence of the wall really does seem to make your catch irrelevant" (Collins 2000, p. 224).

To my mind, the two cases are causally different: Two Catchers is a case of preemptive prevention but Wall is not. However, this is not what this paper aims to establish. Instead, I will examine what implications the two cases have for causation in general *if* they are causally different. In Section 2, I will discuss Collins's account of causal difference, and argue that Collins's account is not adequate. In Section 3, I will argue that the causal difference is not analyzable in terms of either dependence or production. In Section 4, I will examine some non-reductive account of causal difference, and claim that the sense of causal difference relevant to preemptive prevention has yet to be specified.

## 2. Collins's account

Collins proposes that the relevant difference between Two Catchers and Wall is the difference in their backups:

I say that the difference between McDermott's example and mine is this: the counterfactual assumption of the absence of the pure dependence preventer in his story is more far-fetched than the corresponding assumption of absence in mine. It does not require much of a stretch to suppose that I simply get my timing slightly wrong, so that when I leap, I do so at not quite the right moment to be ready to take the catch. It is more far-fetched, on the other hand, to suppose that the brick wall be absent, or that the ball would miraculously pass straight through it. (Collins 2000, p. 229)

A 'pure dependence preventer' is a backup that spoils the counterfactual dependence between a cause and its effect and does so without interfering with any element of the actual sequence (Ibid, p. 227). In each case, the backup spoils the counterfactual dependence between the actual catch and

the absence of the window being broken: without the actual catch, the window would not have been broken due to the backup. In each case, the backup did not interfering with any element of the actual catcher's sequence. Hence, each backup is a pure dependence preventer. On the other hand, the backups, according to Collins, differ in that to suppose that the wall fails to block the ball is too far-fetched, whereas to suppose that the backup catcher fails to catch the ball is not.<sup>1)</sup> On this account, a situation is a case of preemptive prevention only if the counterfactual supposition that the backup fails to be a backup is not far-fetched. In addition, it is a reductive account in that it does not appeal to a causal condition.<sup>2)</sup>

Collins's account is intuitively appealing. It also helps see that a similar type of causal difference can be found in different situations. Consider the following two cases:

Evil Swimmer: Billy sees a child drowning. He thinks he can save the child by jumping into the water, but he does not jump in. The child drowns. Unbeknownst to him, an evil swimmer who does not want Billy to save the child has trodden water on the spot between the shore where Billy stands and the location where the child is struggling. Had Billy jumped into the water,

- 
- 1) In the cited passage, by 'absence', Collins means either non-occurrence such as the brick wall being absent or some distinct occurrence such as the ball passing through the wall. However, to suppose that the wall is absent is as far-fetched as to suppose that the backup catcher is absent. It would be better to suppose that each backup fails to play its role as a backup. Collins's proposal should be understood accordingly.
  - 2) One might wonder why the proposed difference accounts for the causal difference. Possibly, although the backup spoils the counterfactual dependence in both cases, the outcome might counterfactually depend on the actual catch in the case of Two Catchers, but the same does not hold in the case of Wall because the supposition of the absence of the wall is too far-fetched. A related idea will be discussed in detail in the next section.

the evil swimmer would have grabbed him in water and prevented him from saving the child.

Shield: Suzy sees a child drowning. She thinks she can save the child by jumping into the water, but she does not jump in. The child drowns. Unbeknownst to her, there is an impervious, invisible shield between the shore where Suzy stands and the location where the child is struggling. Had Suzy jumped into the water, the shield would have prevented her from saving the child.<sup>3)</sup>

These two cases appear quite similar to each other. However, some differences seem to exist between the two cases. In Evil Swimmer, it is not too far-fetched to suppose that the evil swimmer misses the right timing to grab Billy or is clumsy grabbing him. The outcome, the child's death, does not depend counterfactually on Billy's failure to jump in, thanks to the evil swimmer, but it is not far-fetched to suppose that the evil swimmer fails to grab Billy. On the other hand, in Shield, it is too far-fetched to suppose that the impervious shield suddenly becomes passable or changes its location. In Shield, the outcome does not depend counterfactually on Suzy's failure to jump in thanks to the shield, and it is far-fetched to suppose that the shield fails to block Suzy. If one finds Two Catchers and Wall causally different, she is also likely to find Evil Swimmer and Shield causally different. Collins's account helps to see that. Note that

Evil Swimmer and Shield are putative cases of causation *by* absence, whereas Two Catchers and Wall are putative cases of causation *of* absence (prevention is the causation of absence). If Evil Swimmer and

---

3) The idea of Shield is borrowed from Hall's example of preemptive prevention (Hall 2004, pp. 271-72).

Shield are causal different, they suggest that preemptive prevention is not a causal structure that forms a mere exception that is isolated from the other causal structures.

Collins's account is not adequate, however. To see why, add some details to Wall as follows. Just beneath the ground on which the wall stands (but not on which the building with the window stands), there is an old sewer-pipe that could collapse at any moment. Fortunately, the pipe withstands these conditions at the time of the actual catch and for some duration thereafter. These added details do not appear to make any difference to the initial judgment about Wall that the actual catch did not prevent the window from being broken. With the added details, however, it does not appear far-fetched to suppose that the wall sinks into the ground. Conversely, replace the backup catcher in Two Catchers with a well-made robot nicknamed 'Wall'; it has never missed, and it is 'alert'. Arguably, then, it is far-fetched to suppose that the robot would miss the ball. This, however, does not appear to alter the initial judgment that the actual catch prevented the window from being broken. Two Catchers with this change may be a debatable case, but Wall with the added details is not. Collins's account falls short of adequately capturing the relevant difference.

### 3. Causal difference and two concepts of causation

The concept of causation, according to Hall (2004), is not univocal but comes in two distinct varieties, the production and the dependence concepts. On the production concept, a cause 'generates' or 'produces' its effect. On the dependence concept of causation, an effect counterfactually depends on its cause. The two concepts have different characteristics: on

the production concept, causation is transitive, local or spatiotemporally continuous, and intrinsic, whereas it is not on the dependence concept (Hall 2004, pp. 225-27). Hall agrees with Collins in that cases like *Two Catchers* and *Wall* are causally different. In addition, Hall briefly remarks that these cases “elude classification in terms of production or dependence” (Ibid, p. 271). I claim that there are good reasons to think that these cases are not analyzable in terms of either dependence or production.

Consider the production concept first. On the production concept, causation is a generative or productive relation, and it is transitive, local or spatiotemporally continuous, and intrinsic. As an illustration, consider a familiar case of preemption that does not involve an absence:

Billy-Suzy: Billy throws a rock and it breaks a window. Suzy tries to throw a rock too, but seeing Billy throwing his rock, she does not throw her rock. Had Billy not thrown his rock, Suzy would have thrown her rock and the window would have been broken.

Clearly, Billy’s throw caused the breaking of the window. Intuitively, the process from Billy’s throw to the breaking of the window is spatiotemporally continuous. This is transitive. If Billy’s throw produced the flight of the ball that in turn produced the breaking, then Billy’s throw produced the breaking. And, the process retains its productive relation whether or not a backup like Suzy’s throw is in the vicinity.

In *Two Catchers*, there are intuitively two physical processes: one is the actual catcher’s process and the other is the backup catcher’s process. Similarly, in *Wall*, there are two physical processes: one is the actual catcher’s process and the other is the wall’s process. Intuitively, the actual catcher’s process in *Two Catchers* and the one in *Wall* do not differ

productive-causally: the actual catcher's process in one case (which consists of, say, running toward a certain location, making some preparatory moves in a certain spot, and catching the ball) is quite similar to the actual catcher's process in the other. The backup's process in Two Catchers and the one in Wall do not appear to differ productive-causally either. In Two Catchers, the actual process of the backup consists of, say, running toward a certain location, making some preparatory moves in a certain spot, standing up to walk away and so on. In Wall, the actual process of the backup consists of the wall standing where it is located for the relevant duration of time. There is no noticeable productive-causal difference because there is no productive 'contact' or 'connection' that one case has but the other lacks.

One might argue that consideration of the production concept of causation is a non-starter. Prevention is the causation of an absence, but an absence cannot be productively connected to anything; hence prevention is not a 'genuine' causation on this concept. On the other hand, granting that prevention is not a genuine causation, production theorists agree that prevention is something close to causation (call it 'causation\*'). Consider a situation that is like Two Catchers or Wall but without a backup. In this situation, the actual catch's prevention (that is, the actual catch causing\* the absence) of the window from being broken may be understood in terms of the actual catch of the moving ball (or the interaction between the moving ball's causal process and the actual catcher's causal process) and the following productive-causal counterfactual: without the actual catch, the moving ball would have physically interacted with the window to produce the breaking of the window.<sup>4)</sup>

---

4) Dowe, for example, proposes the following analysis of prevention:



Similarly, the causal\* relation between the actual catch and the window not being broken in Two Catchers may be understood in terms of the actual catch of the moving ball (or the interaction between the moving ball's causal process and the actual catcher's causal process) and the following productive-causal counterfactuals: without the actual catch, the moving ball would have interacted with the backup catcher, and without the actual catch or the backup catch, the moving ball would have interacted with the window to produce the breaking of the window. The same, with some obvious changes, holds for Wall. The causal difference between Two Catchers and Wall is not likely to be captured in productive-causal\* terms either.

What about the dependence concept? Can it help capture the causal difference in question? On the dependence concept of causation, an effect counterfactually depends on its cause, and causation is not transitive, local or spatiotemporally continuous, or intrinsic. Consider again Billy-Suzy. On the dependence concept, causation is not intrinsic. For, whether or not the breaking of the window counterfactually depends on Billy's throw hinges on whether or not the backup is present. On this concept, causation is not transitive either. Consider the counterfactual situation in which Billy does not throw and Suzy throws. The absence of Billy's

---

A prevented B A caused\* not-B if

- (P1) A occurred and B did not, and there occurred an x such that  
 (P2) there is a causal relation between A and the process due to x, such that ...  
 either  
     (i) A is a causal interaction with the causal process x, or  
     (ii) A caused y, a causal interaction with the causal process x,  
 and  
 (P3) if A had not occurred, x would have caused B, where A and B name positive events or facts, and x and y are variables ranging over events and/or facts.  
 (Dowe 2000, p. 132)

throw, intuitively, caused Suzy's throw, and Suzy's throw counterfactually depends on the absence of Billy's throw. Suzy's throw, intuitively, caused the breaking, and given the absence of Billy's throw, the breaking depends counterfactually on Suzy's throw. However, the absence of Billy's throw, i.e., Billy's not throwing his rock, did not cause the breaking. Finally, Billy's throw caused the absence of the window being broken, but it is not connected to the absence via a spatiotemporally continuous process.

The dependence concept of causation in its simplest form, i.e., an effect counterfactually depends on its cause<sup>5</sup>), cannot deal with the preemption regardless of whether it involves an absence or not. For, without an actual cause, its backup would have caused the effect so that the effect does not depend counterfactually on its cause. Moreover, any counterfactual consideration that holds for Two Catchers also appears to hold for Wall: the outcome does not depend counterfactually on the actual catch, the outcome would not have been obtained without either the actual catch or its backup, the outcome would have been obtained with either the actual catch or its backup, and so forth.

On the other hand, the counterfactual dependence between a cause and its effect may be restored by 'controlling' in some way the backup that spoils the counterfactual dependence. For an illustration, consider Billy-Suzy again. What in fact occurred in this case is that Billy threw his ball, Suzy did not, the window was broken, and so forth. Although the breaking of the window does not depend counterfactually on Billy's throw due to the backup (i.e., Suzy's throw), the effect depends on Billy's throw when we hold fixed the fact that Suzy actually did not throw her rock. That is, in the counterfactual situation (closest to the actual world)

---

5) See Lewis (1973). See also Lewis (1986).

where Suzy does not throw her ball as she actually did not but Billy, contrary to fact, does not throw his ball, the window is not broken. On the other hand, holding fixed the fact that Billy actually threw his ball, the breaking of the window does not depend counterfactually on Suzy's throw. Hence, it is Billy's throw that caused the breaking of the window. The dependence concept modified in this way yields the right answer.<sup>6)</sup>

Nevertheless, this modification does not help in preemptive prevention cases. In *Two Catchers*, holding fixed the fact that the backup catcher in fact did not catch the ball, the window would have been broken without the actual catch. The problem is that the same holds for *Wall*. In *Wall*, holding fixed the fact that the wall in fact did not block the ball, the window would have been broken without the actual catch. The two cases do not differ in this regard. Surely, there is something peculiar in the case of *Wall*. To counterfactually suppose that the actual catcher does not catch but also that the wall does not block the ball is, as Collins would point out, far-fetched. Collins's account might help to overcome the difficulty that troubles the modified dependence concept, but it suffers from its own counterexample as shown in the previous section.

One might claim that there is another way to deal with the two cases on the modified dependence concept. Suppose that for *Wall*, the presence of the wall is held fixed, while for *Two Catchers*, the fact that the backup catcher did not catch the ball is held fixed. Holding fixed the fact that the backup catcher did not catch the ball, the window would have been broken without the actual catch. On the other hand, holding fixed the presence of the wall, the window, without the actual catch, would not have been broken thanks to the presence of the wall. Then, the actual

---

6) For related ideas, see Halpern and Pearl (2001), Hitchcock (2001), and Yablo (2002).

catch prevented the window from being broken in Two Catchers, but the actual catch did not in Wall, which is the correct verdict. Similarly, if the fact that the evil swimmer did not interfere with Billy is held fixed in Evil Swimmer, but the presence of the shield is held fixed in Shield, these cases also turn out to be causally different. Therefore, should we treat them differently? This move is ad hoc unless some justification is offered, and it is hard to see why we should not hold fixed the fact that the wall did not block the ball when we should hold fixed the fact that the backup catcher did not catch the ball. I sum, neither the dependence concept nor the production concept is capable of capturing the causal difference.

#### 4. Preemptive prevention and non-reductive accounts of causal difference

Collins's account suggests that our causal judgment is "sensitive to the nature of the backup preventer" (Hall 2004, p. 271), but it has been shown inadequate. There appears to be another sense of 'sensitivity' that is operating in the relevant cases. For example, the wall appears 'insensitive' to other actual goings-on in the situation. The wall is simply 'there' without 'responding to' or 'unimpeded by' the actual catch. Instead, it seems to be 'protected from' or 'unaffected by' other 'influences'. These considerations naturally invite the idea that there might be some causal difference between the backups. That is, in a given situation, the wall persisted for the relevant duration of time, and this sequence of the wall appears to have unfolded as it actually did, not because of the actual catch but *independently of* it. In Two Catchers, however, the actual unfolding of the backup sequence appears 'sensitive

to', 'responding to', 'affected by' or 'impeded by' the other goings-on like the actual catcher's catching the ball. That is, the backup sequence appears to have unfolded as it actually did *because of* the actual catch. The backup sequence, unlike the sequence of the wall, is a sequence whose actual course of unfolding seems to be determined partly by the actual catch.

The above descriptions of the difference between Two Catchers and Wall suggest that a situation is a case of preemptive prevention only if its backup sequence unfolds as it actually does because of the actual preventer. This characterization is different from Collins's account and the two reductive accounts in that although the latter attempts to capture the relevant difference in a non-causal term, i.e., being far-fetched, counterfactual dependence, or physical process, the former characterizes the causal difference in explicitly causal terms. Surely, it is a rough characterization and needs to be refined further.<sup>7)</sup>

This characterization is appealing, though. Consider Evil Swimmer and Shield. In Evil Swimmer, the actual sequence of the evil swimmer appears to have unfolded as it actually did because of Billy's failure to jump in. The evil swimmer has monitored Billy's response to the drowning child, and his subsequent reactions occurred because of Billy's

---

7) In order for this causal account to work, a specific construal of sequence is required: a sequence is not allowed to include an absence. This is because if the backup sequence in Two Catchers includes the backup catcher's not catching the ball, it holds that the backup sequence unfolded as it actually did because of the actual catch: intuitively, the backup catcher did not catch the ball because of the actual catch. Similarly, if the backup sequence in Wall includes the wall not blocking the ball, it holds that the backup sequence unfolded as it actually did because of the actual catch: intuitively, the wall did not block the ball because of the actual catch. If sequence is allowed to include an absence, the alleged causal difference evaporates.

failure to jump in. On the other hand, in Shield, the actual sequence of the shield, i.e., its persisting for some duration, did not react to Suzy's failure to jump in. The actual sequence of the shield does not appear to have unfolded as it actually did because of Suzy's failure to jump in. The characterization also captures the difference in these cases. Could the relevant sense of causation be further specified?

Sartorio (2005) offers the following 'causes as difference-makers principle':

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E. (Sartorio 2005, p. 75).

This principle is not meant to provide a reductive analysis of causation. Instead, it provides a non-reductive account of causal difference-making or "a constraint on theories of causation ... that the true analysis of causation (if there is such a thing) would have to meet" (Ibid). It is non-reductive because it asks us to consider what would not have been *caused* by the absence of cause, not what would have not *occurred* in the absence of cause. Note also that it is only meant to offer a necessary condition for causation. That is, a cause caused its effect only if the effect would not have been caused by the absence of the cause. Sartorio's account is especially relevant to our issue. If it successfully captures the idea of causal difference which has eluded us, this account and the above rough characterization may mutually support each other to reveal the limitation of reductive accounts of causation. Conversely, cases like Two Catchers and Wall may serve as testing cases for the adequacy of this account.<sup>8)</sup>

---

8) For the discussions of the implication of the principle for moral responsibility, see

Consider Two Catchers first. Would the effect, the absence of the window being broken, have been caused by the absence of the actual catch? Clearly, it would not. It is the backup catch that would have caused the effect in the absence of the actual catch. The absence of the actual catch would have caused the backup catcher to catch the ball, and the backup catch would have caused the effect. If causation is transitive, it follows that the absence of the actual catch would have caused the effect. However, intuitively, the absence of the actual catch is an enhancer of the window being broken, not a prohibitor. This strongly suggests that the absence of the actual catch would not have caused the effect, i.e., the absence of the window being broken. Sartorio's account accommodates this intuition and rightly denies the transitivity of causation.<sup>9)</sup> What about Wall? Would the effect have been caused by the absence of the actual catch? Clearly, it would not. The absence of the actual catch is an enhancer of the window being broken, not a prohibitor. In the absence of the actual catch, the effect would have been caused by the wall blocking the ball.

The principle does not capture the causal difference between the two cases since both satisfy Sartorio's principle. The causal difference is not capturable in reductive terms as shown in the previous section, and a promising, non-reductive account of causal difference does not capture it either. This does not mean that the principle should be rejected; it only purports to provide a necessary condition for causation. However, preemptive prevention shows that some other condition has yet to be specified.

---

Sartorio (2013).

9) See Sartorio (2005) for the defense of the claim that "the assumption that we should respect the difference-making idea leads to the rejection of the transitivity of causation" (p. 84).

## 5. Conclusion

Some similar causal structures appear causally different. Putative cases of preemptive prevention offer such an example. However, it is difficult to capture the causal difference. Collins's account is intuitively appealing but inadequate. The causal difference is not analyzable in terms of either dependence or production. A non-reductive account may have a better chance of capturing the causal difference, but the relevant sense of causal difference has yet to be specified. It is well known that any attempt to analyze causation runs into seemingly insurmountable difficulties, and it may well turn out that causation is not analyzable. However, as attested by the discussions of preemptive prevention, the concept of causation, whether it is reducible to something else or not, still requires some further scrutiny.



## References

- Collins, John (2000). Preemptive prevention. *Journal of Philosophy*, 97, 223-34.
- Dowe, Phil (2000). *Physical causation*. Cambridge University Press.
- Hall, Ned (2004). Two concepts of causation. In John Collins, Ned Hall & Laurie Paul (eds.), *Causation and counterfactuals* (pp. 225-276). The MIT Press.
- Halpern, Joseph and Pearl, Judea (2001). Causes and explanations: a structural-model approach part 1: causes. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence* (pp. 194-202).
- Hitchcock, Christopher (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 98, 73-99.
- Lewis, David (1973). Causation. *Journal of Philosophy*, 70, 556-67.
- Lewis, David (1986). Postscripts to 'causation'. In *Philosophical papers*, vol. 2 (pp. 172-213). New York: Oxford University Press.
- McDermott, Michael (1995). Redundant causation. *British Journal for the Philosophy of Science*, 46, 523-44.
- Sartorio, Carolina (2005). Causes as difference-makers. *Philosophical Studies*, 123, 71-96.
- Sartorio, Carolina (2013). Making a difference in a deterministic world. *Philosophical Review*, 122, 189-214.
- Yablo, Stephen (2002). De facto dependence. *Journal of Philosophy*, 99, 130-48.

University of Seoul

Email: [sungsukim@uos.ac.kr](mailto:sungsukim@uos.ac.kr)