Causal Attribution

Causal attributions are central to our ability to make sense of the world and to explain events as well as to plan for the future. Events tend to be preceded by many influences and, in one sense of the word, each is a “cause”; however, in making a judgment, people are selective in choosing only some events as causal. For example, if I slip on the pavement, I may attribute cause to the fact that there was ice-cream on the pavement or to the person that dropped it there. I am unlikely to attribute cause to other events in the causal sequence, to the fact that I bought ice cream or to my decision to follow that particular route.

Our aim is to examine how people decide whether an event caused or prevented a particular outcome. The answer for “cause” and “prevent” may or may not be the same. The statement “c prevented e” may be equivalent in meaning to “c caused not e” as Goldvarg & Johnson-Laird (2001) argue. However, the statements may differ in both their semantics and their pragmatic implications. For instance, people may consider different possibilities in making judgments regarding each (Mandel & Lehman, 1996).

A second theory is based on the idea that a cause involves a process of transmission or exchanges along a causal pathway from cause to effect. This idea can be found in generative theories of causation in philosophy (e.g., Dowe, 2000; Salmon, 1984) and psychology (e.g., Shultz, 1982).

“Make a Difference” Theories of Causation

A number of theories of causation incorporate the notion that a cause is something that makes a difference to the effect, without regard to how the effect is brought about. According to Hume, people infer causation from the regular co-occurrence of distinct events (Hume, 1739/1960) even though this practice cannot ultimately be justified. The great problem of induction according to Hume is that past experience cannot be used to justify prediction without incoherence or circularity. Nevertheless, past experience is all we have and is, therefore, what we use.

More recent proposals about how we know that causes make a difference to their effects include counterfactual and manipulability theories. They also assume that causation can be judged without appealing to the mode of operation of a specific causal mechanism. Counterfactual theories of causation propose that causation can be defined in terms of a counterfactual conditional. In other words, it is the case that “event c caused event e” provided that “if c hadn’t occurred then e wouldn’t have occurred” in the closest possible world to our own (e.g., Lewis, 1973). Using a related idea, manipulability theories propose that “event c causes event e” provided that if I manipulate or intervene in a particular way that involves changing c, then this should change e also (e.g., Halpern & Pearl, 2001; Pearl, 2000).

These approaches capture many of our intuitions, for example, it seems to make sense to say that the alarm not going off caused me to be late, provided that if the alarm had gone off I wouldn’t have been late. Despite this, several major difficulties have emerged for counterfactual theories over the years. One important difficulty is that they can lead to the wrong prediction in cases where the outcome is over-determined, that is, where more than one cause occurred and each alone would have been sufficient to produce the outcome. One of these situations occurs in cases of “pre-emption.” For example, imagine both Billy and Suzy throw
rocks at a bottle and both their throws are right on target. Billy throws first, and his rock hits the bottle and it breaks. In this case, the counterfactual “if Billy hadn’t thrown the rock, then the bottle wouldn’t have broken” is false even though we might say that Billy caused the bottle to break. Lewis has recently reformulated his theory to deal with problems such as this one (Lewis, 2000). He argues that if c causes e if an alteration to c would have led to an alteration of e. If Billy’s rock had been lighter then the shattering of the bottle would have been different, whereas if Suzy’s rock had been lighter, there would be no change to the shattering. Another recent proposal is that people may evaluate cause using counterfactuals that hold certain events constant (Halpern & Pearl, 2001; Hitchcock, 2001; Woodward, 2003). One such counterfactual concerns whether the effect would occur if other causes are assumed to not have occurred. For example, if we maintain that Suzy’s rock didn’t hit the bottle, then if Billy hadn’t thrown, the bottle wouldn’t have broken. It turns out that this is a necessary but not sufficient counterfactual for assessing cause (cf. Woodward, 2003).

Generative Theories of Causation
In contrast to the theories described above, generative theories pay greater attention to the process by which the effect occurs. In this sense, the cause is believed to generate the effect. Causation involves a transmission along a causal pathway (Salmon, 1984; Shultz, 1982) and may involve the exchange of some conserved physical quantity (Dowe, 2000). For example, the reason that we believe that Billy’s rock caused the bottle to break may be that “objecthood” is conserved along the trajectory of Billy’s rock which then transmits a force that breaks the bottle. Suzy’s rock is not the cause of the bottle breaking because no such force is transmitted along the path from Suzy’s throw to the bottle breaking.

Prevention is problematic for generative theories because if A prevents B, then B does not occur and hence there is no continuous process connecting the cause and effect. A possible fix is to assume that prevention is qualitatively different from causation. Dowe (2000) proposed that A prevented B if there was a causal interaction between A and another process x and if A hadn’t occurred then x would have caused B.

Judgments of Causation and Prevention
Despite extensive study, there is no consensus about whether regularity theories or generative theories best describe how people make causal attributions (Baillargeon, Kotovsky & Needham, 1995; Cheng, 1997; Lewis, 2000; Schultz, 1982). Shultz (1982) provided evidence that even young children demand an understanding of a generative transmission process to attribute cause. He showed that children make causal attributions based on inferences about a causal mechanism more readily than based on temporal or spatial contiguity. On the other hand, psychological studies have shown that generating a counterfactual conditional about a possible cause increases attribution to that event (Wells & Gavanski, 1989). Research also suggests that judgments of causation and prevention may rely on different information. For example, judgments of causation may depend on the co-variation of cause and effect, whereas judgments of prevention may depend more on the generation of a counterfactual conditional (Mandel & Lehman, 1996).

Our main aim is to address the question of how people make judgments of both causation and prevention by examining cases that involve a causal mechanism (i.e., a clear process of generative transmission) and ones without a causal mechanism. If people do attribute cause without appealing to a mechanism, then a second aim of our studies was to examine the kinds of counterfactual possibilities that they will consider. For this reason, we used scenarios involving pre-emptive causation.

Experiment 1
We generated two scenarios based on classic examples of late pre-emption in the philosophical literature (cf. Halpern & Pearl, 2001). Each had a similar structure and included a pre-emptive cause. The ‘causation scenario’ included an actual mechanism going from cause to effect, that is, a stone is thrown, it hits a bottle and the bottle breaks:

> There is a bottle on the wall. Billy and Suzy are standing close by with stones and each one throws a stone at the bottle. Their throws are perfectly on target. Billy happens to throw first and his reaches the bottle before Suzy’s. The bottle breaks.

The scenario was followed by two questions:

- Did Billy cause the bottle to break?
- Did Suzy cause the bottle to break?

The ‘prevention scenario’ involved a similar causal structure to the first except that the links were preventive and hence did not involve a mechanism or transfer from cause to effect, that is, a ball is caught and as a result the bottle doesn’t break:

> There is a bottle on the wall. Frank and Jane are standing close by. While they are there someone else aims to throw a ball at the bottle. The aim is perfectly on target. Frank and Jane both step in front of the bottle. Frank happens to step in front of Jane and catches the ball. The bottle doesn’t break.

Again, the scenario was followed by two questions, this time based on prevention:

- Did Frank prevent the bottle from breaking?
- Did Jane prevent the bottle from breaking?

Recent formulations of counterfactual theories developed by Lewis (2000) and Halpern and Pearl (2001) were designed to ascribe causation to Billy and not Suzy. On the assumption that “A prevents B” means the same as “A causes not B”, the theories make the parallel prediction for prevention, ascribing prevention to Frank and not Jane. According to generative mechanism theories, people should also ascribe causation to Billy but not Suzy. In this case, there is a clear mechanism linking the action (Billy’s throw) to the outcome (the bottle breaking). On Dowe’s (2000) account, people should also ascribe prevention to Frank but not Jane, although the reason is different. In this case, the
action (Frank’s catch) interacts with the marble and it is natural to infer that if Frank had not caught the marble, the bottle would have broken. Our aim in this study was to examine whether people’s attributions are consistent with the predictions of these recent theories.

One hundred participants were recruited mainly through a campus-based electronic newspaper and they completed the study on-line. They received both scenarios and responded to the questions by answering “yes” or “no”.

As Table 1 shows, the results for the causation scenario strongly corroborate the predictions of the mechanism and counterfactual views. A large majority (90%) attributed causation to Billy but not Suzy.

For the prevention scenario, the majority of participants also attributed prevention to the first actor only, i.e., Frank but not Jane (60%). The result is consistent with the predictions of recent counterfactual theories and also with Dowe’s account that prevention may involve an interaction with a potential cause.

Table 1: The percentage of “yes” responses to the four questions in experiment 1

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Causation</th>
<th>Prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td>First actor only (Billy / Frank)</td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td>Second actor only (Suzy / Jane)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Both actors</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>Neither</td>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>

However, the percentage of participants who made an attribution to the first actor only was significantly lower in the prevention scenario than in the causation scenario (McNemar Test, p < .001). For the prevention scenario, a large minority (23%) ascribed prevention to both actors. This result is not predicted by Dowe’s mechanism theory or by recent counterfactual theories which would predict that only the first actor prevented the outcome. One explanation for this result is that some people may understand prevention not just as an actual interruption to a causal mechanism (Frank actually caught the ball) but rather as a potential interruption to that mechanism (Jane would have caught the ball). The ball would have transmitted a force breaking the bottle, if Frank and Jane had not intervened to block that process. People often talk about prevention in the sense of having the potential to block some event even if that event does not occur (e.g., the lock is preventing the bike from being stolen).

In our second experiment, we further examine how people understand causation and prevention by asking questions about both in a scenario that involves a causal mechanism and one involving an interruption to a causal mechanism.

**Experiment 2**

We constructed two scenarios based on the ones in the previous experiment. However, in this case, we asked the question about cause and prevention in each case. As before the first scenario contained an actual causal mechanism going from the action to the outcome:

There is a coin standing on its edge at the end of the table. It is unstable and it is about to fall over and land on tails. Billy and Suzy are standing close by with marbles. Each one rolls their marble down the table towards the coin. Their rolls are perfectly on target and each one will hit the coin at exactly the same spot, knock it the other way, and the coin will land on heads. Billy happens to roll first and his marble reaches the coin before Suzy’s. The coin falls over and lands on heads.

After reading the scenario, participants answered the following four questions. Half of them answered the cause questions first and half answered the prevention questions first.

Did Billy cause the coin to land on heads?
Did Suzy cause the coin to land on heads?
Did Billy prevent the coin from landing on tails?
Did Suzy prevent the coin from landing on tails?

The second scenario was similar but it contained no actual causal mechanism going from the action to the effect:

There is a coin standing on its edge at the end of the table. It is unstable and it is about to fall over and land on heads. Frank and Jane are standing close by with marbles. While they are there someone else rolls a marble toward the coin. The roll is perfectly on target and it will hit the coin, knock it over and the coin will land on tails. Frank and Jane both reach out and put their hands in front of the coin. Frank happens to put his hand in front of Jane’s and he catches the marble. The coin falls over and lands on heads.

After reading the scenario, participants answered the same questions as after the ‘mechanism’ scenario. Again, half of them answered the cause questions first and half answered the prevention questions first.

Did Frank cause the coin to land on heads?
Did Jane cause the coin to land on heads?
Did Frank prevent the coin from landing on tails?
Did Jane prevent the coin from landing on tails?

Recent counterfactual theories make the same prediction for both scenarios. They predict that people should attribute causation and prevention to the first actor only. According to Dowe’s mechanism theory people should attribute causation when there is a mechanism producing the outcome but not when the mechanism is blocked. In contrast, they should attribute prevention when the mechanism is blocked but not when it produces the outcome.
Hence, people should make attributions of causation but not prevention to Billy in the mechanism scenario and attributions of prevention but not causation to Frank in the mechanism-blocked scenario.

Fifty seven participants were again recruited through a campus-based electronic newspaper and completed the study on-line. They read both scenarios and in each case, they responded to each of the four questions by answering "yes" or "no". The order of presentation of the scenarios and the order of cause and prevention questions were randomized.

As Table 2 shows, the majority of participants attributed causation to the first actor only in the mechanism scenario where there was a transmission along the pathway from cause (Billy throws a marble) to effect (the coin lands on heads; 74%). However, in the mechanism-blocked scenario when there was no continuous process from cause (Frank catches the marble) to effect (the coin lands on heads), the majority of participants attributed causation to neither (77%). The percentage of participants who made an attribution to the first actor only was significantly lower in the mechanism scenario than in the mechanism-blocked scenario (McNemar Test, p < .001). The results support the predictions of mechanism theories that causal judgments depend on the presence of a causal mechanism.

In contrast, the pattern of prevention attributions was similar for the mechanism and mechanism-blocked scenarios. Approximately half of the participants attributed prevention to the first actor only (54% and 52% for the mechanism and no-mechanism scenarios respectively, and these did not differ significantly, p > .9). These judgments conform to the predictions of recent counterfactual theorists.

Table 2: The percentage of “yes” responses to the eight questions in experiment 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>First actor only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Billy/Frank</td>
<td>74</td>
<td>54</td>
<td>18</td>
<td>52</td>
</tr>
<tr>
<td>Second actor only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suzy/Jane</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Both actors</td>
<td>10</td>
<td>12</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Neither</td>
<td>16</td>
<td>30</td>
<td>77</td>
<td>34</td>
</tr>
</tbody>
</table>

In contrast to the previous experiment, where the largest minority attributed prevention to both actors, in this case approximately one third of participants attributed prevention to neither actor regardless of whether a mechanism was present (30%) or not (34%). The main difference between the two studies is that in the previous study the prevention scenario ended with no change of state, the bottle remained unbroken, whereas in this experiment, the scenarios end with an outcome different from the original state, i.e., the coin was spinning but in the end lands on heads. In this case, the outcome may be attributed to a different mechanism, namely the spinning of the coin, and hence the actors may be perceived to have played a lesser role and are thus less likely to be assigned any kind of causal role.

One alternative explanation for the results is that attributions of causation depend not on the presence of a mechanism but rather on a change to the default outcome.1 In the mechanism scenario, the marble hitting the coin changes the outcome from tails to heads. In contrast in the mechanism-blocked scenario, the marble doesn’t hit the coin and as a result, there is no change to the direction of the coin’s fall. We carried out our third experiment to test this explanation.

### Experiment 3

In our third experiment, we aimed to test whether causation depends on a causal mechanism or merely a change to the outcome. For this reason, we constructed a scenario without a causal mechanism linking the action to the outcome but one which did involve a change to the final outcome. The scenario is similar to the mechanism-blocked scenario used in Experiment 2 but rather than blocking the marble, the action involved lifting a book which was blocking the path to the coin:

> There is a coin wobbling on edge at the end of the table. It is about to fall over and land on tails. There is a book directly in front of the coin. Max and Anne are standing close by. While they are there someone else rolls a marble toward the coin. The roll is perfectly on target and in the absence of the book it will hit the coin, knock it over and the coin will land on heads. Max and Anne both reach out to lift the book. Max happens to reach in front of Anne and he lifts the book. The marble hits the coin, and the coin falls over and lands on heads.

After reading the scenario, participants answered the following four questions. Half of them answered the cause questions first and half answered the prevention questions first.

> Did Max cause the coin to land on heads?
> Did Anne cause the coin to land on heads?
> Did Max prevent the coin from landing on tails?
> Did Anne prevent the coin from landing on tails?

If attribution of cause depends on a change in the outcome and not on a causal mechanism, then we expect people to attribute cause to the first actor (Max) in this scenario. In contrast, if a mechanism from cause to effect is important in attributing cause, then we expect that people will again tend not to attribute cause to Max.

---

1 We thank Jim Woodward for suggesting this explanation.
Sixty-eight participants (recruited as in the previous experiments) read the scenario and responded to each of the four questions by answering “yes” or “no”.

**Table 3: The percentage of “yes” responses to the eight questions in experiment 3**

<table>
<thead>
<tr>
<th></th>
<th>Mechanism unblocked Change in the Outcome</th>
<th>Mechanism unblocked Change in the Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cause question</td>
<td>Prevent question</td>
</tr>
<tr>
<td>First actor only (Max)</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td>Second actor only (Anne)</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Both actors</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Neither</td>
<td>62</td>
<td>41</td>
</tr>
</tbody>
</table>

As Table 3 shows, the majority of participants did not attribute causation to either actor (62%). The result supports the predictions made by mechanism theories. Even when there is a change to the default outcome, in the absence of a mechanism, people are less likely to make a causal inference.

In real life, most cases of prevention do not involve a change to the outcome. It is perhaps unsurprising then that the pattern of results for the prevention question was quite close to that found in Experiment 2. Approximately half of the participants judged the first actor only to have prevented the outcome (47%) and more participants judged the first actor to have prevented than to have caused the outcome (McNemar, Test, p < .04). Again as in Experiment 2, a large minority believed that neither participant prevented the outcome (41%). The results suggest that a change in the outcome is not important for prevention judgments.

**General Discussion**

The results of our experiments suggest that cause and prevent may not have symmetric meanings. Judgments of whether A caused B tend to depend on the presence of a causal mechanism. In contrast, for the majority of people, judgments of prevention conform to the predictions of recent counterfactual theories (Halpern & Pearl, 2001; Lewis, 2000). But in certain cases, prevention may be attributed by virtue of a potential interruption to a causal mechanism.

In our first experiment, we compared judgments on two scenarios involving pre-emption; in one an action produced an outcome and in the other an action prevented an outcome. The results showed that people made different attributions in each case. When the action produced the outcome, there was a clear causal mechanism connecting the cause (the stone is thrown) to the outcome (the bottle is hit and breaks) and people view the action as causal in this case. Also, when the action interrupted the mechanism, the majority attributed prevention to the first actor but not the actor who was pre-empted. The results corroborate both Dowe’s mechanism and recent counterfactual theories. However, a large minority viewed both actors as preventors, suggesting that they understand prevention to mean a potential interruption to a causal mechanism.

The results of our second experiment suggested that both mechanisms and counterfactuals are important. We constructed two scenarios, one with a causal mechanism connecting the cause to the effect and one in which a causal mechanism is blocked, and we asked causation and prevention questions about each. The results showed that causation questions were strongly sensitive to the presence of a causal mechanism corroborating the predictions of mechanism theories but not counterfactual theories. Participants attributed a cause to the actor only when the scenario involved a causal mechanism. In contrast, many participants attributed prevention without regard for whether a mechanism was present or was blocked. In the second experiment, fewer attributions were made to both actors, perhaps because an actual outcome was produced by another mechanism in this case.

In our third experiment, we ruled out an alternative explanation for our results, namely that attributions of causation depend not on a causal mechanism but rather on a change to the outcome. We constructed a scenario in which no mechanism linked the action to the outcome but the action facilitated a mechanism that did change the outcome. The majority of participants judged that the action did not cause the outcome supporting the view that it is a mechanism rather than a change to the outcome that is necessary for attributions of causation. Overall, the results suggest that an understanding of causal mechanisms as well as how people generate counterfactuals is necessary for a complete theory of the meaning of causation and prevention.

The results suggest further questions for psychological research. First, we need a greater understanding of how people reason with different types of causal mechanism. The examples used in our experiments all involved physical relations. But many causal relations are not of this nature. For example, causation can involve social influence or people may have reasons which lead them to carry out a particular action. Furthermore, physical mechanisms can involve a transmission process involving obvious physical contact between cause and effect as in our examples of the marble hitting the coin, or the transmission process may be invisible if wind from a fan caused the coin to fall (Shultz, 1982). In some cases the mechanism may be unknown.
Indeed, the very notion of mechanism requires further elaboration. Schaffer (2000) offers multiple examples of causes that disconnect a cause from its effect. For instance, pulling a trigger causes a gun to shoot through disconnection: It moves a part (the sear) that otherwise would inhibit a spring from uncoiling, and the action of the spring causes an explosion that propels the bullet. For a mechanism theory to be viable, it must define people’s understanding of mechanism to include such cases.

Our results suggest that judgments of causation may be highly sensitive to the kinds of questions that people are asked (Hilton, 1990). People may make different judgments depending on whether they are to decide if something caused or enabled an outcome (Goldvarg & Johnson-Laird, 2001) or when they are asked to make social judgments of blame or responsibility. The results may also vary depending on the framing of the question. For example, we may respond differently when asked if something is “the cause” or “a cause”.

The task we posed to participants was essentially linguistic, asking them whether the verb “cause” or “prevent” was an appropriate characterization of a scene. We, like everyone else, assume that such linguistic judgments derive from a conception of the scene. If we are right that the notion of causal mechanism is necessary to explain how people attribute cause, then that suggests that people have access to a notion of mechanism that could be critical in a variety of other conceptual tasks as well, like explanation, induction, and decision making (cf. Sloman, 2005).

Causal judgments are ubiquitous in our everyday thinking as well as in domains ranging from science to the law. We suggest some steps toward the development of an understanding of this process.

Acknowledgments

We thank Denis Hilton, Joe Halpern, Rui da Silva Neves and Jim Woodward for discussions on this topic. This research was supported by NASA grant NCC2-1217 to Steven Sloman.

References