When Possibility Informs Reality: Counterfactual Thinking as a Cue to Causality

Barbara A. Spellman[1] and David R. Mandel

Department of Psychology, University of Virginia, Charlottesville, Virginia (B.A.S.), and Department of Psychology, University of Hertfordshire, Hatfield, Hertfordshire, England (D.R.M.)

Abstract

People often engage in counterfactual thinking, that is, imagining alternatives to the real world and mentally playing out the consequences. Yet the counterfactuals people tend to imagine are a small subset of those that could possibly be imagined. There is some debate as to the relation between counterfactual thinking and causal beliefs. Some researchers argue that counterfactual thinking is the key to causal judgments; current research suggests, however, that the relation is rather complex. When people think about counterfactuals, they focus on ways to prevent bad or uncommon outcomes; when people think about causes, they focus on things that covary with outcomes. Counterfactual thinking may affect causality judgments by changing beliefs about the probabilities of possible alternatives to what actually happened, thereby changing beliefs as to whether a cause and effect actually covary. The way in which counterfactual thinking affects causal attributions may have practical consequences for mental health and the legal system.

Keywords

counterfactuals; causality; reasoning

> If I *were* in New York right now, and if I happened to be standing on a sidewalk on my way to lunch, waiting for the light to change, and if a car happened to jump the curb, I might be struck dead. By not being there, I may have freed that space on the sidewalk for someone else who might be standing there and get run over. And in that event you might say that I'm partially responsible for that death. In a weak sense I'd be responsible at the end of a long causal chain. We're all linked by these causal chains to everyone around us. . . .
>
> McInerney (1992, p. 120)

Most of us have outlandish suppositional "what if" thoughts like this from time to time. We may wonder, "What if my parents had never met?" (*Back to the Future*) or "What if I had never been born?" (*It's a Wonderful Life*). Seemingly less outlandish and far more ubiquitous are our thoughts about the ways in which things might have been different "if only" an event leading up to a particular outcome had been different. For instance, suppose that you get into a car accident while taking a scenic route home that you rarely use but chose to take today because it was particularly sunny. If you are like many other people, you might "undo" the accident by thinking something like "If only I had taken my usual route, I wouldn't have been in the accident" (Kahneman & Tversky, 1982; Mandel & Lehman, 1996). These "if only" thoughts are termed *counterfactual conditionals* because they focus on a changed (or *mutated*) version of reality (i.e., one that is counter to fact) and because they represent a conditional if-then relation between an antecedent event (e.g., the route taken) and a consequent event (e.g., getting into an accident).

CAUSAL CONSEQUENCES OF COUNTERFACTUAL THINKING

Counterfactual thinking has consequences beyond mere daydreams or entertainment. Engaging in counterfactual thinking may influence people's causal attributions.

Practical consequences of this relation can be seen for mental health and the legal system.

For example, counterfactual thinking can produce what Kahneman and Miller (1986) termed *emotional amplification*—a heightening of affect brought about by realizing that an outcome was not inevitable because it easily could have been undone. Counterfactual thinking can amplify feelings of regret, distress and self-blame, and shame and guilt, as well as satisfaction and happiness (see Roese & Olson, 1995, for reviews). Perhaps recurrent counterfactual thinking, in which people obsessively mutate their own actions, leads people to exaggerate their own causal role for both their misfortunes and their successes, resulting in those heightened emotions.

Counterfactual thinking can also affect liability and guilt judgments in legal settings. For example, in one study, subjects read a story about a date rape and then listened to a mock lawyer's closing argument suggesting possible mutations to the story (Branscombe, Owen, Garstka, & Coleman, 1996). If the argument mutated the defendant's actions so that the rape would be undone, the rapist was assigned more fault (cause, blame, and responsibility) than if his actions were mutated but the rape still would have occurred. Similarly, when the victim's actions were mutated, she was assigned more fault if the mutation would undo the rape than if the rape still would have occurred.

THE METHOD BEHIND OUR MUSINGS

Although in our counterfactual musings there are an infinite number of ways we can mutate antecedents so that an outcome would be different, we tend to introduce relatively minor mutations that are systematically constrained (see Roese, 1997). People commonly mutate abnormal events by restoring them to their normal default (Kahneman & Miller, 1986). For example, in the car-accident story, people undo the accident by restoring the abnormal antecedent of taking the unusual route to its normal default—taking the usual route (Kahneman & Tversky, 1982). People also tend to mutate antecedents that are personally controllable (see Roese, 1997). For instance, one is more likely to think, "If only I had decided to take the usual route . . ." than to think, "If only it wasn't sunny" or even "If only the other driver was not reckless" (Mandel & Lehman, 1996). More generally, research on counterfactual thinking demonstrates that people entertain a very small and nonrandom sample from the universe of possible mutations.

Just as various antecedent events leading up to an outcome are differentially likely to be mutated, various outcomes are differentially likely to trigger counterfactual thinking. Negative events and abnormal events are the most likely triggers (see Roese, 1997; Roese & Olson, 1995). Consequently, research on counterfactual thinking has focused on subjects' counterfactual thoughts about negative events like car accidents, food poisonings, and stock-market losses, and unexpected positive events like winning a lottery.

COUNTERFACTUAL THINKING AS TESTS OF CAUSALITY

Kahneman and Tversky's (1982) highly influential chapter on the *simulation heuristic* described the basic logic by which if-only thinking could be used to explore the causes of past events. They proposed that people often run an if-then simulation in their minds to make various sorts of judgments, including predictions of future events and causal explanations of past events. Accordingly, to understand whether X caused Y, a person may imagine that X had not occurred and then imagine what events would follow. The easier it is to imagine that Y would not now follow, the more likely the person is to view X as a cause of Y.

Kahneman and Tversky's (1982) proposal is certainly plausible: They were not suggesting that all counterfactual-conditional thinking is necessarily directed at understanding the causes of past events, but simply making the point that such thinking could be useful toward that end. Nor did they argue that all forms of causal reasoning are equally likely to rely on counterfactual thinking; their examples suggested that counterfactual thinking would be most useful for understanding the causes of a single case (as in law) rather than for understanding causal relations governing a set of events (as in science). Indeed, their proposal echoed earlier work on causal reasoning by the philosopher Mackie (1974) and the legal philosophers Hart and Honoré (1959/1985), and amounted to what in political science and law is assumed generally to be an important criterion for attributing causality: In order to ascribe causal status to an actor or event, one should believe that *but for* the causal candidate in question, the outcome would not have occurred. In other words, one must demonstrate that the causal candidate was *necessary*, under the circumstances, for the outcome to occur. Despite the modest scope of Kahneman and Tversky's proposal, however, many attribution researchers subsequently posited that counterfactual thinking plays a primary role in causal attributions.

EMPIRICAL EVIDENCE

The first empirical study to measure both counterfactual thinking and judgments of causality in the same story (Wells & Gavanski, 1989) supported the idea that mutability was the key to causality. Subjects read about a woman who was taken to dinner by her boss. The boss ordered for both of them, but the dish he ordered contained wine—to which the woman was allergic. She ate the food and died. The boss had considered ordering something else: In one version, the alternative dish also contained wine; in the other version, the alternative dish did not contain wine. Subjects were asked to list four things that could have been different in the story to prevent the woman's death (mutation task) and to rate how much of a causal role the boss's ordering decision played in her death. Relative to subjects who read that both dishes contained wine, subjects who read that the other dish did not contain wine (a) were more likely to mutate the boss's decision and (b) rated the boss's decision as more causal. These results were interpreted as showing two things: that for causality to be assigned to an event (the boss's decision), the event must have a counterfactual that would have undone the outcome (i.e., mutability is necessary for causality) and that the more available a counterfactual alternative is, the more causal that event seems (i.e., ease of mutability affects perceived causal impact).

Later research, however, has found different patterns for mutability and causality in other, richer, scenarios. For example, in another study (Mandel & Lehman, 1996), subjects read a story about a car accident like the one alluded to earlier (based on Kahneman & Tversky, 1982): On a sunny day, Mr. Jones decided to take an unusual route home; being indecisive, he braked hard to stop at a yellow light; he began driving when the light turned green; he was hit by a drunk teenager who charged through the red light. One group of subjects was asked how Jones would finish the thought "If only . . . ," and another group was asked what Jones would think caused the accident. The subjects tended to mutate Jones's decision to take the unusual route and his indecisive driving but to assign causality to the teenager's drunk driving. Hence, the most mutated event need not be perceived as the most causal event.

Therefore, although sometimes mutability and causality converge (as in the food decision), sometimes they differ (as in the car accident). Yet the latter experiment does not disconfirm that mutability is necessary for causality: The drunk driver's actions—the most causal event—can easily be mutated to undo the outcome (even though other aspects of the story are mutated more often).

However, there is a situation that demonstrates that mutability is *not necessary* for causality. Suppose that Able and Baker shoot Smith at the exact same instant. Smith dies and the coroner reports that either shot alone would have killed him (i.e., there is *causal overdetermination*). Able can argue that he did not kill Smith because even if he had not shot Smith, Smith would have died anyway. Mutating Able's actions alone does not work; in fact, subjects mutate Able's and Baker's actions together to undo Smith's death. But when attributing causality, subjects judge Able and Baker individually to be causes of Smith's death and sentence Able and Baker each to the maximum jail time (Spellman, Maris, & Wynn, 1999). Both philosophers and legal theorists consider such cases of causal overdetermination to be an exception to the but-for causality rule. Because causality is assigned even when mutation does not undo the outcome, mutability is not necessary for causality.

Moreover, mutability is *not sufficient* for causality. For example, one might believe that Jones's car accident would not have occurred if it had not been a sunny day (as Jones would have taken his normal route); however, no one would argue that the sunny day caused the accident. Obviously, therefore, not everything that can be mutated to undo an outcome is judged as a cause of that outcome.

RECENT THEORETICAL PERSPECTIVES

The relation between mutability and causality is not simply one of identity, necessity, or sufficiency; rather, a more complex picture has emerged.

According to a recent proposal (Mandel & Lehman, 1996), everyday counterfactuals are not tests of causality. Rather, for negative outcomes, counterfactuals are after-the-fact realizations of ways that would have been sufficient to prevent the outcomes—and especially ways that actors themselves could have prevented their misfortunes. Data from the car-accident experiment reported earlier support this account. In addition to the counterfactual- and causal-question groups, a third group of subjects was included in this experiment. These subjects answered a question about how Jones would think the accident could have been prevented. Not surprisingly, subjects' answers for the preventability question focused on the same aspects of the story as the answers for the counterfactual question, but were different from the answers for the causality question. Thus, al-

though, logically, sufficient-but-foregone preventors (e.g., taking the usual route) seem like necessary causes, psychologically, attributions of preventability and causality are different. In attributing preventability, people focus on controllable antecedents (e.g., choice of route, stopping at a yellow light); in attributing causality, people focus on antecedents that general knowledge suggests would covary with, and therefore predict, the outcome (e.g., drunk drivers).

The *crediting causality* model (Spellman, 1997) tells us how a covariation analysis of causal candidates might be performed in reasoning about single cases. According to this model, causality attributions in single cases are analogous to causality attributions in science. In science, to find whether something is causal (e.g., whether drunk driving causes accidents), people presumably compare the probability of the effect when the cause is present (e.g., the probability of accidents given drunk drivers) with the probability of the effect when the cause is absent (e.g., the probability of accidents given no drunk drivers); if the former probability is larger, people infer causality. For the single case, people may consider and compare that probability with the probability of an accident before that event occurred (i.e., the baseline probability). If the event (e.g., a drunk driver) raises the probability of the outcome, then it is causal; if it does not (e.g., a sunny day), it is not.

Counterfactual thinking might affect causal attributions by acting as input to that probability comparison. If one can counterfactually imagine many alternatives to an outcome like an accident, then the baseline probability of the accident seems low, and, therefore, the causality of the event in question (e.g., drunk driver) seems high. If there are no alternatives to an outcome (e.g., all foods contain wine and will cause illness), then the baseline probability of the outcome is high, and the causality of the event in question (e.g., ordering a particular dish) seems low. Similarly, in the experiment involving counterfactual thinking about the date rape (Branscombe et al., 1996), thinking about mutations that would undo the rape may have changed estimates of the baseline probability of a rape, and thus affected the subjects' legal judgments. Note that when it seems that nothing can be mutated to undo the outcome, people feel the outcome was inevitable; that is, it was due to fate (Mandel & Lehman, 1996).

CONCLUDING REMARKS

The philosopher David Hume described causation—or, more accurately, causal thinking—as the "cement of the universe" because it binds together people's perceptions of events that would otherwise appear unrelated. It seems, however, that our beliefs about the universe of the actual, to which Hume referred, are affected by our considerations of the merely possible—created by the "what ifs" and "if onlys" of counterfactual thinking.

Recommended Reading

Kahneman, D., & Miller, D.T. (1986). (See References)
Mackie, J.L. (1974). (See References)
Mandel, D.R., & Lehman, D.R. (1996). (See References)
Roese, N.J. (1999). *Counterfactual Research News* [On-line]. Available: http://www.psych.nwu.edu/psych/people/faculty/roese/research/cf/cfnews.htm
Roese, N.J., & Olson, J.M. (Eds.). (1995). (See References)

Acknowledgments—This work was supported in part by a NIMH FIRST Award to the first author and by Research Grant No. 95-3054 from the Natural Sciences and Engineering Research Council of Canada to the second author. We thank Darrin Lehman and Angeline Lillard for helpful comments on an earlier draft.

Note

1. Address correspondence to Barbara A. Spellman, Department of Psychology, 102 Gilmer Hall, University of Virginia, Charlottesville, VA 22903; e-mail: spellman@ virginia.edu.

References

Branscombe, N.R., Owen, S., Garstka, T.A., & Coleman, J. (1996). Rape and accident counterfactuals: Who might have done otherwise and would it have changed the outcome? *Journal of Applied Social Psychology, 26,* 1042–1067.
Hart, H.L., & Honoré, A.M. (1985). *Causation in the law* (2nd ed.). Oxford, England: Oxford University Press. (Original work published 1959)
Kahneman, D., & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.
Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.
Mackie, J.L. (1974). *The cement of the universe: A study of causation.* Oxford, England: Oxford University Press.
Mandel, D.R., & Lehman, D.R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology, 71,* 450–463.
McInerney, J. (1992). *Brightness falls.* New York: Knopf.
Roese, N.J. (1997). Counterfactual thinking. *Psychological Bulletin, 121,* 133–148.
Roese, N.J., & Olson, J.M. (Eds.). (1995). *What might have been: The social psychology of counterfactual thinking.* Mahwah, NJ: Erlbaum.
Spellman, B.A. (1997). Crediting causality. *Journal of Experimental Psychology: General, 126,* 323–348.
Spellman, B.A., Maris, J.R., & Wynn, J.E. (1999). *Causality without mutability.* Unpublished manuscript, University of Virginia, Charlottesville.
Wells, G.L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56,* 161–169.