

# THE RELATION BETWEEN COUNTERFACTUAL (“BUT FOR”) AND CAUSAL REASONING: EXPERIMENTAL FINDINGS AND IMPLICATIONS FOR JURORS’ DECISIONS

BARBARA A. SPELLMAN\* AND ALEXANDRA KINCANNON\*\*

## I

### INTRODUCTION

People often try to discover the causes of the events around them: Why did their child behave a certain way? Why did their football team lose? Why does their computer crash at critical moments? People also often trust other people to make causal decisions for them. We allow scientists to tell us that smoking causes lung cancer and that eggs are bad for us (some years, anyway). We also allow, or rather we *require*, jurors to make decisions about causality in many kinds of cases.

Another kind of reasoning people do is to imagine the world other than it is and play out the consequences. For example, if the child had not watched so much television, if the quarterback had not thrown that interception, or if grandfather had quit smoking, might the world be a happier place? This kind of reasoning is called “counterfactual reasoning.” In its most common guise, people imagine an early event (“antecedent”) as being different (for example, no interception), leading to an outcome (“consequent”) that may also be different (for example, winning the game).

#### A. Counterfactual Reasoning as “but for” Causal Reasoning in the Law

The legal system often asks jurors to use counterfactual reasoning to make decisions about causation. Obviously, the idea of causality is essential in law: Generally, we do not wish to punish or impose liability on someone unless he or she caused the damage at issue. Both criminal and civil codes (especially in the context of negligence) spend many pages defining what it means to be a cause. In both kinds of cases, to be considered a cause, the potential cause in question must fulfill two requirements: It must be a “but for” cause (also called “cause-

---

Copyright © 2001 by Barbara A. Spellman and Alexandra Kincannon

This article is also available at <http://www.law.duke.edu/journals/64LCPSpellman>.

\* Associate Professor of Psychology, University of Virginia. This research was supported by an NIMH Grant to the author.

\*\* Ph.D. Candidate (anticipated May 2002), University of Virginia.

in-fact” or “factual cause” or “sine qua non”) of the outcome; and it must be a legal cause (also called “proximate cause”) of the outcome. The legal cause requirement limits the infinite number of “but for” causes to those for which it makes sense to hold people liable.

The definition of “but for” cause can be seen in the Model Penal Code (“MPC”), the Restatement (Second) of Torts, and in common civil jury instructions. In the MPC’s general definition of “causal relationship,” the “but for” requirement is stated as follows:

Conduct is the cause of a result when: (a) it is an antecedent but for which the result in question would not have occurred. . .<sup>1</sup>

In negligence cases in California, the Book of Approved Jury Instructions (“BAJI”) recommended the following “standard” jury instruction for many years:

A proximate cause of injury is a cause which . . . produces the injury and without which the injury would not have occurred.<sup>2</sup>

Thus, jurors’ causal judgments require counterfactual reasoning and a finding that changing the antecedent would have changed the outcome in question.

## B. Present Questions

The legal system obviously assumes that people can reason counterfactually by making this type of reasoning essential for judgments of causation. What does the psychological literature say about people’s capacity for counterfactual and causal reasoning and the relation between them?

This article describes the existing body of psychological research on counterfactual and causal reasoning. It then focuses on research dealing with two specific questions regarding their relation that is relevant to legal causal judgments. First, is counterfactual reasoning really the basis for causal judgments? We show how psychology’s answer to that question has changed over time, and present research involving cases of multiple sufficient causes that demonstrates that counterfactual reasoning cannot, in fact, be the basis for all causal judgments.

---

1. MODEL PENAL CODE § 2.03(1) (1985).

2. COMMITTEE ON STANDARD JURY INSTRUCTIONS, CALIFORNIA JURY INSTRUCTIONS CIVIL, BOOK OF APPROVED JURY INSTRUCTIONS (BAJI) § 3.75 (7th ed. 1986 & Supp. 1994). The full text of BAJI § 3.75 defining proximate cause is below. The judge selects which of the bracketed terms to use:

A proximate cause of [injury] [damage] [loss] [or] [harm] is a cause which, in natural and continuous sequence, produces the [injury] [damage] [loss] [or] [harm] and without which the [injury] [damage] [loss] [or] [harm] would not have occurred.

*Id.*

The 8th edition of BAJI § 3.76 has similar bracketed language. See COMMITTEE ON STANDARD JURY INSTRUCTIONS, BOOK OF APPROVED JURY INSTRUCTIONS (BAJI) § 3.76 (8th ed. 1995).

Second, in what way does counterfactual reasoning affect causal reasoning? We review research that shows that making a counterfactual judgment can affect later causal judgments; we then make suggestions as to when and why it does so—and should do so. Finally, we consider the implications of this research for jurors' causal judgments.

## II

### STUDYING CAUSAL AND COUNTERFACTUAL REASONING IN THE LABORATORY

Much of what psychologists know about how people make causal and counterfactual judgments derives from studies in which research subjects (usually, but not always, college students recruited for either money or course credit) come into the laboratory, read a vignette or scenario, and then respond to directed questions.<sup>3</sup>

#### A. Counterfactual Reasoning

Studies of counterfactual reasoning have been conducted primarily by social psychologists. In typical studies, subjects read a story involving several events that might have contributed to some (usually bad) outcome; car accidents, muggings, and food poisonings are favorites. For example, in a famous study we will return to later,<sup>4</sup> subjects read about a man, Mr. Jones, who chose to drive home by a scenic route that he rarely used, because it was a particularly sunny day. He braked hard to stop at a yellow light, but when he began to cross after the light turned green, a truck driven by a teenage boy on drugs charged into the intersection and struck the car, killing Mr. Jones. After reading such a story, subjects may be asked to make a direct counterfactual reasoning judgment in one of two ways. They may be asked to complete an “if only” statement. For example, in this study they were asked: “As commonly happens in such situations, the Jones family and their friends often thought and often said, ‘if only . . . ,’ during the days that followed the accident. How did they continue the thought? Please write one or more likely completions.”<sup>5</sup> Alternatively, subjects may be asked to list some number of ways in which the story could be changed so that the outcome would be different or “undone.” Researchers assume that the event written down first by the subject is what the subject thought

---

3. A few researchers have studied counterfactual and causal reasoning about “real” events. See, e.g., Christopher G. Davis et al., *Self-Blame Following a Traumatic Event: The Role of Perceived Avoidability*, 22 PERSONALITY & SOC. PSYCHOL. BULL. 557 (1996) (discussing counterfactual reasoning about one's own injury after accident); Christopher G. Davis et al., *The Undoing of Traumatic Life Events*, 21 PERSONALITY & SOC. PSYCHOL. BULL. 109 (1995) (discussing counterfactual reasoning after the loss of spouse or child); Victoria Husted Medvec et al., *When Less is More: Counterfactual Thinking and Satisfaction Among Olympic Medalists*, 69 J. PERSONALITY & SOC. PSYCHOL. 603 (1995); Bernard Weiner, “Spontaneous” Causal Thinking, 97 PSYCHOL. BULL. 74 (1985) (reviewing spontaneous causal attributions).

4. See Daniel Kahneman & Amos Tversky, *The Simulation Heuristic*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 204-06 (Daniel Kahneman et al. eds., 1982).

5. *Id.* at 204.

was the most “mutable” event—that is, the one easiest to imagine changing.<sup>6</sup> Note that in the “if only” procedure, the events that subjects write down need not be events that actually would change the outcome; for example, a subject could write “If only I had been nicer to him.” Yet researchers have found that overwhelmingly, subjects will complete the “if only” statement with a counterfactual that would undo the outcome.<sup>7</sup>

In addition to these direct assessments of counterfactual reasoning by the “if only” and “undoing” questions, counterfactual reasoning is sometimes assumed and then assessed indirectly. For example, subjects might read one of two versions of a story in which a person goes to a usual or unusual restaurant or takes a usual or unusual route home from a bar and ends up sick or injured.<sup>8</sup> Subjects then answer questions about how much sympathy they feel for the protagonist or how much compensation they would award the protagonist. Typically, subjects are more sympathetic and award more compensation when the protagonist takes the unusual action—presumably because it is easier to imagine a counterfactual alternative action that would not have led to the bad outcome (that is, doing the usual thing).

Neal J. Roese distinguishes between factors that activate counterfactual reasoning and factors that influence the content of what will be mutated in counterfactual reasoning.<sup>9</sup> Because jurors are already in a context in which they are asked to do such reasoning, it is the latter that is of more interest in this article. In general, studies suggest: (1) that exceptional or unusual events will be mutated more often than normal or usual ones; (2) that events having to do with the focus of the story or questions will be mutated more often than those that do not; (3) that controllable events will be mutated more often than uncontrollable ones; and (4) that immoral events will be mutated often more than moral events.<sup>10</sup> Note, however, that most of these findings are not without exception.

In addition, some researchers have argued that earlier events in a story are more mutable than later ones;<sup>11</sup> others have argued that later events in a story

---

6. See, e.g., Ahogni N'gbala & Nyla R. Branscombe, *Mental Simulation and Causal Attribution: When Simulating an Event Does Not Affect Fault Assignment*, 31 J. EXPERIMENTAL SOC. PSYCHOL. 139 (1995); Gary L. Wells et al., *The Undoing of Scenarios*, 53 J. PERSONALITY & SOC. PSYCHOL. 421 (1987). For an opinion that directly challenges this assumption, see Barbara A. Spellman, *The Relation Between Counterfactual and Causal Reasoning: Some Methodological and Theoretical Issues* (2000) (unpublished manuscript, on file with author).

7. See Nyla R. Branscombe et al., *Rape and Accident Counterfactuals: Who Might Have Done Otherwise and Would It Have Changed the Outcome?*, 26 J. APPLIED SOC. PSYCHOL. 1042 (1996).

8. See, e.g., C. Neil Macrae, *A Tale of Two Curries: Counterfactual Thinking and Accident-Related Judgments*, 18 PERSONALITY & SOC. PSYCHOL. BULL. 84 (1992); C. Neil Macrae & Alan B. Milne, *A Curry for Your Thoughts: Empathic Effects on Counterfactual Thinking*, 18 PERSONALITY & SOC. PSYCHOL. BULL. 625 (1992); C. Neil Macrae et al., *Counterfactual Thinking and the Perception of Criminal Behaviour*, 84 BRIT. J. PSYCHOL. 221 (1993).

9. See Neal J. Roese, *Counterfactual Thinking*, 121 PSYCHOL. BULL. 133, 137-39 (1997).

10. See *id.* at 137 (reviewing the studies).

11. See, e.g., N'gbala & Branscombe, *supra* note 6; Gary L. Wells et al., *supra* note 6; 53 J. PERSONALITY & SOC. PSYCHOL. 421 (1987).

are more mutable than early ones.<sup>12</sup> Dale T. Miller and Saku Gunasegaram argue that the relative mutability of events may depend on the relation between the events: If the events form a causal chain in which the earlier events cause the later ones, then the first will be more mutable; if the events form a temporal chain in which the events are independent, then the last will be more mutable.<sup>13</sup>

## B. Causal Reasoning

Causal reasoning has been studied by psychologists in many specialty areas including animal, developmental, cognitive, and social. As a result, several different kinds of experimental tasks have been used. We distinguish two kinds of methodologies used to study causal reasoning;<sup>14</sup> the literatures have remained fairly distinct. The methodologies vary on the kind and quantity of information given to the subjects. In some experiments, the subjects are asked to make judgments that seem scientific; in others, the judgments seem legal.

In the multiple case scientist/probabilistic approach, subjects learn about the covariation between possible causes and effects. Information is presented either in a table giving a summary of the number of times the cause and effect occur separately and together, or in a series of presentations in which the subject experiences (for example, by watching) the relative frequency of the combinations of cause and effect. For example, subjects may see many examples of pictures of a fertilizer being poured on a plant (or not) and then see whether the plant bloomed. Subjects are usually asked to rate the efficacy of the cause on a rating scale, or to make a probability judgment about how often the effect will occur in the presence of the cause in the future. One theory is that causal

---

12. See, e.g., Dale T. Miller & Saku Gunasegaram, *Temporal Order and the Perceived Mutability of Events: Implications for Blame Assignment*, 59 J. PERSONALITY & SOC. PSYCHOL. 1111 (1990).

13. See *id.* at 1116. But see the discussion of order effects in causal judgments below for another interpretation.

14. There is historically a very important third methodology that is not relevant to the present discussion. We call it the "multiple case social/covariation approach." Whereas Methodology 1 involves repeated exposures to the same stimuli over and over, this methodology usually involves learning information about how different people react to different stimuli on different occasions. This research is driven by Harold H. Kelley's ANOVA model of causal induction. See Harold H. Kelley, *Attribution Theory in Social Psychology*, 15 NEB. SYMP. MOTIVATION 192 (1967). Subjects might learn, for example, that all of our papers are accepted on the first try at *Psychological Review*; that none of our colleagues' papers are accepted on the first try at *Psychological Review*; and that none of our papers are accepted on the first try at any other journal. They are then asked why a particular paper was accepted: Was it something about the person, the stimulus, the situation, or some interaction that caused the outcome? See Patricia W. Cheng & Laura R. Novick, *Covariation in Natural Causal Induction*, 99 PSYCHOL. REV. 365 (1992) (arguing that their probabilistic contrast model which, on the surface, computes contingencies of the form of Methodology 1, is actually an extension of Kelley's ANOVA model).

Factors that affect these causal judgments include completeness of information and background knowledge. See Patricia W. Cheng & Laura R. Novick, *A Probabilistic Contrast Model of Causal Induction*, 58 J. PERSONALITY & SOC. PSYCHOL. 545 (1990), and the context or framing of the causal inquiry. See Patricia W. Cheng & Laura R. Novick, *Causes Versus Enabling Conditions*, 40 COGNITION 83 (1991); Ann L. McGill, *Context Effects in Judgments of Causation*, 57 J. PERSONALITY & SOC. PSYCHOL. 189 (1989).

judgments should depend on the *contingency* between the cause and effect.<sup>15</sup> Contingencies are computed as shown below: by taking the proportion of times the effect occurs in the presence of the cause and subtracting the proportion of times the effect occurs in the absence of the cause. In other words, if a possible cause increases the probability of the effect, relative to the baseline occurrence of the effect, we attribute causality.<sup>16</sup>

$$(1) \text{ causal efficacy (contingency)} \rightarrow \text{proportion of times effect occurs in the presence of the cause} - \text{proportion of times effect occurs in the absence of the cause (baseline)}$$

Factors that affect these causal judgments include (1) the probability of the effect occurring in the presence of the cause and the probability of the effect occurring in the absence of the cause (that is, the contingency); (2) the overall prevalence of the effect; (3) the presence of alternative causal factors; and (4) a theory of why a factor might be causal.<sup>17</sup>

The single case “legal” approach methodology is the one most similar to that used by researchers studying counterfactual reasoning. In the single case approach to studying causal reasoning, subjects read a story involving several events that might have contributed to some (usually bad) outcome. The stories can be identical to the stories used to study counterfactual reasoning. After reading the story, subjects typically are asked to make a causal judgment. Such judgments tend to take the form of either listing (subjects are asked to list the causes of the outcome, like the procedure for counterfactual reasoning studies), rating (subjects are given a list of potential causes and are asked to rate each one of them as to causality on some rating scale), or both (subjects are asked to list then rate causes). The judgments are not always about causality per se. Although subjects are often asked to list or rate “causes,” they may be asked to make judgments about responsibility, blame, control, foreseeability, monetary damages, or jail sentences, all of which have been interpreted as being relevant to causal judgments.<sup>18</sup>

Many of the factors that influence counterfactual reasoning have also been shown to affect this type of causal reasoning, including normality,<sup>19</sup> controllabil-

---

15. See Herbert M. Jenkins & William C. Ward, *Judgments of Contingency Between Responses and Outcomes*, 79 PSYCHOL. MONOGRAPHS: GENERAL & APPLIED 1 (1965).

16. See Marc J. Buehner & Patricia W. Cheng, *Causal Induction: The Power PC Theory Versus the Rescorla-Wagner Model*, in PROCEEDINGS OF THE NINETEENTH ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY 55-60 (1997); Patricia W. Cheng, *From Covariation to Causation: A Causal Power Theory*, 104 PSYCHOL. REV. 367 (1997) (arguing persuasively that a better method would take into account the base rate of the effect).

17. See Cheng, *supra* note 16.

18. Of course, we know that these factors are not identical. See generally KELLY G. SHAVER, *THE ATTRIBUTION OF BLAME: CAUSALITY, RESPONSIBILITY AND BLAMEWORTHINESS* (1985).

19. See Denis J. Hilton & Ben R. Slugoski, *Knowledge-Based Causal Attribution: The Abnormal Conditions Focus Model*, 93 PSYCHOL. REV. 75 (1986).

ity,<sup>20</sup> and morality.<sup>21</sup> Temporal order is also an issue: Some researchers have found that early events are more causal than later ones;<sup>22</sup> other researchers have found that later events are more causal than earlier ones.<sup>23</sup> One of the authors, in uniting the two approaches, however, has shown that the effect of temporal order is minimal; what really matters is how much each event contributes to the outcome.<sup>24</sup>

Barbara Spellman has argued that, even in single-case kinds of situations, causality is attributed analogously to Methodology 1; that is, people take into account the extent to which a cause changes the probability of the outcome.<sup>25</sup> Spellman's subjects read about two players, Allen and Barry, who played as a team in the final round of a game show. Allen had to try to complete a simple task (fill a water balloon). His speed determined which game Barry would then have to play to win the prize. In some conditions, Barry's two potential games were of the same difficulty; in others, the game Barry would play if Allen were fast would be easier than the game he would play if Allen were slow. Subjects were told the probabilities of each player winning each game. Subjects did not always attribute more causality to the first or last player; rather, they seemed to take into account how much each player changed the probability of the outcome. Spellman argued that, when assigning causality to causes that act in sequence, people might consider the following:<sup>26</sup>

$$(2) \text{ causal contribution} \rightarrow \text{probability of the effect after the cause occurs} - \text{probability of the effect before the cause occurs (baseline)}$$

Note that the process is very similar to that for making contingency judgments (Methodology 1), in that causality is assigned based on how much a cause raises the probability of the effect compared to the baseline probability of the effect occurring in the absence of the cause.

### III

#### IS COUNTERFACTUAL REASONING THE BASIS FOR CAUSAL JUDGMENTS?

Early researchers in counterfactual reasoning took the view that counterfactual reasoning was at the heart of causal reasoning; the current view is that they

20. See N'gbala & Branscombe, *supra* note 6.

21. See Mark D. Alicke, *Culpable Causation*, 63 J. PERSONALITY & SOC. PSYCHOL. 368 (1992); N'gbala & Branscombe, *supra* note 6.

22. See Joel T. Johnson et al., *Causal Primacy and Comparative Fault: The Effect of Position in a Causal Chain on Judgments of Legal Responsibility*, 15 PERSONALITY & SOC. PSYCHOL. BULL. 161 (1989); Amiram Vinokur & Icek Ajzen, *Relative Importance of Prior and Immediate Events: A Causal Primacy Effect*, 42 J. PERSONALITY & SOC. PSYCHOL. 820 (1982).

23. See N'gbala & Branscombe, *supra* note 6.

24. See Barbara A. Spellman, *Crediting Causality*, 126 J. EXPERIMENTAL PSYCHOL.: GENERAL 323 (1997).

25. See *id.*

26. See *id.* at 332-40.

are different and that, at a minimum, they answer different questions and serve different functions.<sup>27</sup> Below we describe the justifications for each of those beliefs.

#### A. Early View: Counterfactual Reasoning is at the Heart of Causal Reasoning

The notion that counterfactual reasoning might be at the heart of causal reasoning has both philosophical and psychological underpinnings. Daniel Kahneman and Amos Tversky's highly influential article on the "simulation heuristic"—the source of the car accident story described above—describes the way in which counterfactual thinking can be used to generate causal attributions.<sup>28</sup> 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. Therefore, to understand whether an antecedent (for example, taking an unusual route home) caused an outcome (for example, the car accident), we may imagine that the antecedent had *not* occurred and then imagine what events would follow. The easier it is to imagine that the outcome would *not* now follow, the more likely we are to view the antecedent as a cause of the outcome. Kahneman and Tversky did not claim that causal reasoning had to be done that way, merely that counterfactual thinking could be useful toward that end.<sup>29</sup>

*1. Philosophical Basis.* Kahneman and Tversky's suggestion is thus a proposal that "but for" counterfactual reasoning, as required by the legal definitions above, could be psychologically important when making everyday causal judgments.<sup>30</sup> In fact, many researchers have referred to the work of the philosopher John Mackie<sup>31</sup> and the legal philosophers H.L.A. Hart and Tony Honoré<sup>32</sup> when trying to characterize the psychology of causality.

For example, in her review article, Marlys G. Lipe argues that all social psychological theories of causal attribution are based on counterfactual reasoning.<sup>33</sup> She suggests that when an event occurs for which we want to know the cause (for example, the shattering of a glass), and we think of a possible cause (for example, the opera singer hitting a high note), we test the causal status by asking a

---

27. See David R. Mandel & Darrin R. Lehman, *Counterfactual Thinking and Ascriptions of Cause and Preventability*, 71 J. PERSONALITY & SOC. PSYCHOL. 450 (1996); Neal J. Roese & James M. Olson, *Functions of Counterfactual Thinking*, in WHAT MIGHT HAVE BEEN: THE SOCIAL PSYCHOLOGY OF COUNTERFACTUAL THINKING 169 (Neal J. Roese & James M. Olson eds., 1995).

28. Kahneman & Tversky, *supra* note 4, at 201-08.

29. See *id.* at 202-03.

30. See *id.*

31. JOHN L. MACKIE, *THE CEMENT OF THE UNIVERSE: A STUDY OF CAUSATION* (1974).

32. See H.L.A. HART & TONY HONORÉ, *CAUSATION IN THE LAW* 1 (2d ed. 1985) (arguing that the notion of what counts as a cause under the law is the same as our notion of cause in everyday life and that it is the "plain man's notions of causation (and not the philosopher's or the scientist's) with which the law is concerned").

33. See generally Marlys Gascho Lipe, *Counterfactual Reasoning as a Framework for Attribution Theories*, 109 PSYCHOL. BULL. 456 (1991).

counterfactual question of the form: “Would the effect have occurred if the proposed cause had not?” Lipe relies on Mackie for support and says:

Mackie . . . argued that our concept of causation is intimately tied to such counterfactual questions. He argued that when we are able to imagine or observe instances of the effect (the shattering of the goblet) without the proposed cause (the high note), causality is not inferred. However, when we are unable to imagine such alternative situations, the proposed causal link remains intact. Hart and Honoré . . . , cited in Mackie . . . , stated that the lawyer approaches causal statements in the following way, “When it is suggested that A is the cause of B he is apt to ask as the first question, would B have happened without A?”<sup>34</sup>

Thus, psychologists seem to be relying on legal philosophers to understand how people actually reason about causality.

2. *Empirical Support.* What empirical support was there for the view that counterfactual reasoning was necessary for causality? Early empirical support for the belief came from a study in which subjects read a story about a woman who was taken out to dinner by her boss.<sup>35</sup> The boss orders dinner for both of them, but the dish contains an ingredient—wine—to which the woman is allergic. She eats the dish, gets sick, and dies. In both conditions of the experiment the boss had considered ordering something else: In one condition (one-wine), the other dish did not contain the fatal ingredient; in the other condition (two-wine), the other dish also contained 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 decision to order that dish played in her death. Subjects in the one-wine condition—where the alternative choice did not contain wine—were more willing to mutate the selection of the dish and rated the boss’s decision as more causal. These researchers argued that counterfactual reasoning affected causal judgments because in the scenario in which there was a counterfactual alternative to the boss’s decision that would have undone the outcome, the boss was viewed as more causal (and therefore that the availability of the alternative affected that causal judgment).<sup>36</sup> In fact, it was so easily assumed that counterfactual reasoning *was* causal reasoning in that, several experiments around 1990, researchers asked subjects to make counterfactual reasoning judgments and subsequently made claims about causal reasoning judgments.<sup>37</sup> According to such researchers, “[t]he general proposition guiding this research is that an event will

---

34. *Id.* at 457-58 (citations omitted).

35. See Gary L. Wells & Igor Gavanski, *Mental Simulations of Causality*, 56 J. PERSONALITY & SOC. PSYCHOL. 161 (1989).

36. However, it is not clear why that relationship is implied by the results. There was a change made to the story and that change might have affected counterfactual reasoning and causal reasoning separately and directly without the former mediating the latter.

37. See, e.g., Igor Gavanski & Gary L. Wells, *Counterfactual Processing of Normal and Exceptional Events*, 25 J. EXPERIMENTAL SOC. PSYCHOL. 314 (1989); Vittorio Girotto et al., *Event Controllability in Counterfactual Thinking*, 78 ACTA PSYCHOLOGICA 111 (1991).

be judged as causal of an outcome to the extent that mutations to that event would undo the outcome.”<sup>38</sup>

#### B. Later View: Counterfactual Reasoning is not at the Heart of Causal Reasoning

Subsequent research that used dependent measures of both mutability and causality has revealed that counterfactual and causal judgments need not be congruent, and that the assumed relation between them therefore does not hold. For example, Ahogni N’gbala and Nyla R. Branscombe asked subjects to read a story in which a boy was injured in an automobile accident after the following sequence of events: (1) his father failed to pick him up from school; (2) a neighbor gave him a ride; and (3) a drunk driver crashed into the neighbor’s car.<sup>39</sup> Subjects first did a mutability task that required them to complete the sentence stem: “The outcome might have been different IF ONLY . . .” Subjects then rated each story character as to causality for the outcome. Judgments of mutability and causality differed: Although most subjects mutated the father’s actions, more causality was assigned to the drunk driver’s actions. Thus, the most mutated event need not be the most causal event. Note, however, that in their stimuli, the event rated as the most causal (driver’s actions) was clearly a “but for” cause of the outcome.

Similar results were found by David R. Mandel and Darrin R. Lehman using a variation of Kahneman and Tversky’s car accident story described above.<sup>40</sup> They asked one group of subjects how Jones would finish the thought “If only . . .” and another group what Jones would think caused the accident. The subjects tended to mutate Jones’s decision to take the unusual route but to assign causality to the teenager’s drunk driving. Hence, the most mutated event need not be perceived as the most causal event. In addition, a third group of subjects was asked what Jones would think about when he considered how the accident could have been prevented. The pattern of answers in the preventability condition was similar to that of the counterfactual condition.

Mandel and Lehman interpret these results by arguing that mutability and causality ask different questions, and that people use different reasoning processes when deciding what is mutable and what is causal. They argue that, when deciding what to mutate, people focus on things that will prevent the outcome, whereas, when deciding what is the cause, people focus on events that co-vary with the outcome.<sup>41</sup>

Note an important common feature in all of the experiments described so far: The events that subjects list as most causal are also “but for” causes; that is,

---

38. Wells & Gavanski, *supra* note 35, at 161.

39. See N’gbala & Branscombe, *supra* note 6.

40. In this version Jones is injured but does not die.

41. See *id.*; see also Spellman, *supra* note 6 (containing a series of experiments in which events that are the most often mutated are not the events rated the most causal, and demonstrating that, whereas it is easy to affect what subjects will write down as things they would mutate, it is harder to affect what subjects will rate as causal).

being a “but for” or necessary cause is a requirement for assigning causality. In fact, Mandel and Lehman cautiously state that they do *not* “deny claims that causes must be mutable.”<sup>42</sup> None of the above studies, however, has looked at cases of multiple sufficient causes—that is, cases where it seems causality should be attributed to events that are *not* “but for” causes.

### C. Multiple Sufficient Causes: Distinguishing Counterfactual and Causal Reasoning

Consider the following fact pattern:

Reed hates Smith and wants to kill him. West also hates Smith (for an entirely different reason) and also wants to kill him. One day Reed shoots Smith in the head. At the exact same instant, West shoots Smith in the heart. Smith dies. The coroner says that either shot alone would have been enough to kill Smith.

Who should be punished for Smith’s death? When Reed goes to trial he argues that he should not be punished because he did not cause Smith’s death: Even if he had not shot Smith, Smith would still be dead. West, of course, would argue exactly the same thing. The result of this analysis is that both should go free<sup>43</sup>—clearly not an outcome that would please the general public!

This classic “conundrum of causation”<sup>44</sup> illustrates the problem of assigning causality in cases involving “multiple sufficient causes” or “causal overdetermination”—situations in which two or more causes (for example, two gunshots) are each sufficient for bringing about the effect (for example, the victim’s death). These situations pose a problem for philosophers, lawyers, and psychologists. The essence of the problem lies in how we determine causality. Under the usual “but for” definition of a cause, both Reed and West are not causes and must be allowed to go free.

*1. Experimental Illustrations.* How do uninstructed laypeople attribute causality in the case of multiple sufficient causes? In the following two experiments, we contrast the situations of multiple *sufficient* causes with multiple *necessary* causes.<sup>45</sup> Multiple sufficient causes are illustrated by the initial example above. Smith is killed by two simultaneous gunshots—either shot alone would have been enough to kill him, but neither alone is necessary. To change the scenario into one of multiple necessary causes, imagine that the coroner had said that either shot alone would *not* have been enough to kill Smith. It is only

42. Mandel & Lehman, *supra* note 27, at 460.

43. See Robert N. Strassfeld, *If . . . : Counterfactuals in the Law*, 60 GEO. WASH. L. REV. 339, 354 (1992).

44. LEO KATZ, *BAD ACTS AND GUILTY MINDS: CONUNDRUMS OF THE CRIMINAL LAW* (1987) (noting other interesting examples).

45. These results are from Barbara A. Spellman et al., *Causal Attribution in Cases of Causal Overdetermination: Reasoning When the Legal “But For” Test Fails* (2000) (unpublished manuscript, on file with author); see also Barbara A. Spellman, *The Construction of Causal Explanations* (1993) (unpublished doctoral dissertation, on file with author) (noting similar results).

because *both* happened that he died. In the latter version, both shots are necessary but neither alone is sufficient.

Note the difference between the results of mutating Reed and West in the two gunshot scenarios. In the original sufficient story, mutating Reed's actions individually does not change the outcome of Smith dying. In the later necessary story, it does change the outcome of Smith dying. Those who believe that "but for" reasoning is the basis for causality also believe that an action that, when undone, undoes the harm, will be viewed as more causal than an action that, when undone, does not undo the harm.<sup>46</sup> Thus, such a theory predicts that subjects who read the necessary story should attribute more causality to Reed for Smith's death than those who read the sufficient story, because in the former, changing Reed's actions does change the outcome, whereas in the latter it does not. In fact, the strong "but for" view would argue that, in cases of multiple sufficient causes, neither cause should be rated as causal at all.

The present experiments were designed to determine whether subjects use "but for" reasoning to make causal attributions in cases of simultaneous multiple sufficient causes in the absence of any specific instruction as to how to assess causality.

#### *a. The Double Murder*

*i. Method.* College student subjects read variations of the Reed and West murder story presented above. Subjects in the sufficient condition read that exact story. Subjects in the necessary condition read the same facts except that the coroner said that either gunshot alone would not have killed Smith. Thus, in the sufficient condition mutating either Reed or West alone would not change the outcome of Smith's death; rather, both would have to be mutated together. However, in the necessary condition, mutating either person's actions alone would change the outcome of Smith's death.

Half the subjects in each condition were asked to mutate events in the stories so that the outcome (Smith's death) would be different. Specifically, they were told to list four ways in which the story could be changed so that the outcome would be different. The other half were asked to reason about the causes of the outcome. First, they listed as many factors as they thought caused the outcome, and then they rated the importance of each factor listed with regard to causality on a scale of zero (not causal at all) to ten (total cause). All subjects were then asked to sentence Reed and West to jail time. They were told that in the state in which Smith was shot, the maximum jail sentence for murder is fifty years, and they were asked how much jail time, from a minimum of zero to a maximum of fifty years, they would give each of the defendants.

---

46. See Gavanski & Wells, *supra* note 37; Richard L. Wiener & Christine C. Pritchard, *Negligence Law and Mental Mutation: A Social Inference Model of Apportioning Fault*, in *APPLICATIONS OF HEURISTICS AND BIASES TO SOCIAL ISSUES* 117 (L. Heath et al. eds., 1994).

*ii. Results.* The first question to ask is whose actions did subjects in the mutability condition mutate? As one would expect, the subjects' answers depended on whether they read the sufficient or the necessary scenario. Subjects in the necessary condition were happy to mutate either Reed's or West's actions individually (both together = 14%, either Reed or West alone = 71%) whereas subjects in the sufficient condition realized that they had to mutate Reed's and West's actions together in order to change the outcome (both together = sixty-eight percent, either Reed or West alone = twenty-three percent).

The second question is, who did subjects in the causality condition list as causal? The answers did not depend on the scenario. In each condition, subjects tended to list either Reed or West *individually* as a cause of Smith's death for their first response, rather than listing both together in their first response.

The key questions concern causal ratings and jail time. Given that mutating Reed or West alone does not change the outcome (they are not "but for" causes) in the sufficient scenario, are they rated as causal at all? Do subjects think they should be given jail time? In fact, contrary to what "but for" reasoning would predict, subjects rated Reed and West as significantly more causal in the sufficient scenario (mean = 8.2), when mutating their actions wouldn't change the outcome, than in the necessary scenario (mean = 5.7), when mutating their actions would change the outcome. Similarly, subjects who read the sufficient scenario (mean number of years = 46.3) assigned more jail time than subjects who read the necessary scenario (mean number of years = 43.8).

*b. Multiple sufficient causes without intent: The Storm.* The previous experiment has been criticized for incorporating a problematic factor—Reed and West's *intent* to kill Smith. Perhaps subjects consider Reed and West both to be causes because they think the characters are bad people and deserve to be punished. To get around this problem, we devised another multiple sufficient cause story that eliminates human action.

*i. Method.* Subjects read one of two variations of a story about a storm. In both versions they read that a textile warehouse had caught fire and burned to the ground. Film from a security camera showed that at the same time: (1) a bolt of lightning had struck one side of the warehouse, starting a fire; and (2) fierce winds blew an electrical utility pole down into the other side of the building, resulting in a large shower of electrical sparks that started another fire. Subjects who read the sufficient condition learned that either fire alone would have been enough to burn down the entire building. Subjects who read the necessary condition learned that each fire alone would have burned down only half the building.

Again, half the subjects in each condition were asked to mutate events in the stories so that the outcome would be different. The other half were asked to list and rate the causes of the outcome.

*ii. Results.* The results of the storm experiment were similar to those of the double murder experiment. In the mutability condition, which subjects were more likely to mutate depended on the scenario they read. Subjects in the necessary condition were more likely to mutate either the wind or lightning alone (both together = 28%; either alone = 50%) whereas subjects in the sufficient condition realized that they had to mutate both the wind and the lightning in order to change the outcome (both together = 46%; either alone = 17%). In the causality condition, all subjects were more likely to list either the wind or the lightning alone as causal rather than both together.

Furthermore, subjects rated the wind and lightning as significantly more causal in the sufficient scenario (mean = 7.0), when mutating them individually wouldn't change the outcome, than in the necessary scenario (mean = 5.9), when mutating them individually would change the outcome.

*c. Summary.* These two experiments examined the relationship between mutability and causality to discover whether mutability, or "but for" causality, is essential for attributing causality. Our conclusion is that subjects are not using "but for" reasoning to attribute causality in these cases. Subjects are quite willing to attribute causality to events that, when mutated, do not change the outcome. In the causal-sufficient condition, they list both antecedents (both of the shooters or both the lightning and the wind) as causal and give causality ratings substantially above zero (and substantially above those in the necessary condition) to both antecedents in the story. If subjects were using "but for" causality, neither sufficient antecedent should be listed or rated as causal at all.

*2. Reactions to the Failure of "But For" Reasoning.* If "but for" reasoning does not work in cases of multiple sufficient causes—and by "does not work" we mean neither gets us to the just result, nor reflects what people actually do—how should the legal system and psychological theorists respond? The problems with the definition of causality raised by such examples have been of both practical and theoretical concern. This interest can be seen not only in the extensive discussions by legal philosophers such as Hart & Honore<sup>47</sup> and legal commentators such as William L. Prosser,<sup>48</sup> but also in a plethora of law review articles.<sup>49</sup>

*a. The exception approach.* One view is to argue that we should use "but for" reasoning, *except* in cases of multiple sufficient causes, when another ap-

---

47. See HART & HONORÉ, *supra* note 32.

48. See WILLIAM L. PROSSER, *HANDBOOK OF THE LAW OF TORTS* (4th ed. 1971).

49. See Eric Amsel et al., *Do Lawyers Reason Differently From Psychologists? A Comparative Design for Studying Expertise*, in *COMPLEX PROBLEM SOLVING: PRINCIPLES AND MECHANISMS* 223, 226 (Robert J. Sternberg & Peter A. Frensch eds., 1991); W. PAGE KEETON ET AL., *PROSSER AND KEETON ON THE LAW OF TORTS* 263 (5th ed. 1984) [hereinafter *PROSSER & KEETON*] (listing several review articles).

proach should be used to assign causality. This solution is neither theoretically parsimonious nor efficient for instructing juries.

Both the drafters of the Model Penal Code and the California Committee on Jury Instructions recognized the shortcomings of “but for”-type language in cases involving multiple sufficient causes. They responded to these shortcomings in different ways. The comments to the Model Penal Code address exactly the problem of two gunshots. They state that it was the drafters’ intent to make both parties liable. Rather than add language to cover such cases, the comments state that, in cases of multiple sufficient causes, “the result in question should be viewed as including the precise way in which the forbidden consequence occurs.”<sup>50</sup> In the gunshot story, the result in question would be death from *two* gunshot wounds. Thus, if either Reed or West had not shot Smith, the outcome would have been different; it would have been death from *one* gunshot wound.

The California Committee acknowledged that in cases of multiple sufficient causes BAJI 3.75 was not helpful, and suggested that a different instruction be given in such cases.<sup>51</sup>

BAJI 3.78 Concurring Causes—Either Cause Alone Sufficient

Where two causes combine to bring about an injury and either one of them operating alone would have been sufficient to cause the injury, either cause is considered to be a proximate cause of the injury if it is a material element and a substantial factor in bringing it about, even though the result would have occurred without it.<sup>52</sup>

Such instructions allow jurors to decide that both Reed and West are causes of Smith’s death. Failure to give this instruction, rather than the previous one, in appropriate cases, would result in reversible error.<sup>53</sup>

Thus, the Model Penal Code uses the “but for” test with multiple sufficient causes, but addresses the problem by changing the definition of what counts as the “result,” whereas the California instructions added specific language addressing the “exception” of multiple sufficient causes to the “but for” rule.<sup>54</sup>

A third way of addressing the problem of applying the “but for” causality rule is to drop the rule entirely. In fact, the California Committee withdrew its approval of both BAJI 3.75 and BAJI 3.78, and now recommends a “substantial factor” test, which can apply to both cases involving multiple sufficient causes and the more usual cases of those that do not:

The law defines cause in its own particular way.<sup>55</sup> A cause of injury is something that is a substantial factor in bringing about an injury.

---

50. MODEL PENAL CODE AND COMMENTARIES 259 (1985).

51. COMMITTEE ON STANDARD JURY INSTRUCTIONS, *supra* note 2, § 3.78.

52. *Id.*

53. *See id.*

54. *See* RESTATEMENT (SECOND) OF TORTS §§ 431-32. The Restatement view is somewhat unclear; it seems to require “but for” causality to qualify as a substantial factor except in cases of multiple sufficient causes. *See* Strassfeld, *supra* note 43, at n.73.

55. COMMITTEE ON STANDARD JURY INSTRUCTIONS, *supra* note 2, § 3.76.

Substantial factor tests have been adopted by many states. Such tests, however, also have their own set of problems.<sup>56</sup>

b. *The “don’t worry, this never happens” approach.* When we initially submitted our multiple sufficient cause experiments to a not-to-be-named journal, one reviewer made the comment that “[t]he kinds of cases in which ‘but for’ reasoning comes into conflict with multiple sufficient cause reasoning, particularly those that go to jury trial, are vanishingly rare. They are the stuff of the first week of Law School and the works of Leo Katz.”<sup>57</sup>

As noted before, however, even if such cases rarely happen, they have garnered more than a bit of concern from the legal community. Moreover, cases of multiple causes—perhaps not simultaneous, perhaps not sufficient—are not as rare as that reviewer suspects. And even if they are rare, there might be a common psychological mechanism behind reasoning in all of them.

The present experiments both involve *simultaneous* multiple sufficient causes (for example, the two guns were shot at the same time, or the two fires began at the same time). Recently, Erich J. Greene and John M. Darley reported results using scenarios involving *sequential* multiple sufficient causes.<sup>58</sup> In their basic story, Harold decides to kill Joe using a poison pill. He sneaks the pill into Joe’s vitamin bottle, and Joe takes it and dies. The pill is both necessary and sufficient for Joe’s death. In numerous variations of the story, the pill is either not necessary (because Joe is killed in a car accident or by a gunshot before he is killed by the poison) or not sufficient (because the dose is not strong enough). Subjects were asked to judge the sufficiency and necessity of Harold’s conduct in contributing to Joe’s death, the contribution of Harold’s conduct in contributing to Joe’s death, and Harold’s liability in Joe’s death. They discovered, *inter alia*, that subjects would attribute liability even when “but for” tests of causality fail, and that although necessity and sufficiency were good predictors of liability, the best unique predictor of liability was judged to be contribution. Thus, these data involving *sequential* multiple sufficient causes also argue against “but for” causality and support Spellman’s<sup>59</sup> crediting causality hypothesis.

In fact, there is a class of real-life cases that involve questions of multiple causes that may be both sufficient and simultaneous. For example, there are cases involving multiple simultaneous polluters of a stream or people stricken with emphysema due, allegedly, to exposure by asbestos manufactured by many different companies.<sup>60</sup> Four New York City police officers were tried for homi-

---

56. See PROSSER & KEETON, *supra* note 49, at 240; Strassfeld, *supra* note 43, at 353-57.

57. Anonymous review from a not-to-be-named journal (Jan. 7, 1999).

58. See Erich J. Greene & John M. Darley, *Effects of Necessary, Sufficient, and Indirect Causation on Judgments of Criminal Liability*, 22 LAW & HUM. BEHAV. 429 (1998); see also PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* 181-89 (1995).

59. See Spellman, *supra* note 24, at 323-48.

60. See Gerald W. Boston, *Apportionment of Harm in Tort Law: A Proposed Restatement*, 21 DAYTON L. REV. 267 (1996) (providing actual examples).

cide after simultaneously shooting a man, Amadou Diallo, forty-one times. In their article in this volume, Lawrence M. Solan and John M. Darley discuss other real examples of potential multiple sufficient causes.<sup>61</sup>

#### IV

##### HOW COUNTERFACTUAL REASONING INFLUENCES CAUSAL REASONING

###### A. Demonstrations of the Effect

It has been argued that the mere existence of a counterfactual alternative that would change an outcome will affect causal reasoning.<sup>62</sup> The effect of actually thinking about counterfactual alternatives has been demonstrated in experiments that either allow subjects to think freely about them or that direct their thoughts to specific counterfactual alternatives.<sup>63</sup> Subjects in the Branscombe et al. studies read stories about either a date rape or car accident. In one experiment, some subjects were asked to read a story about a rape and then to change the behavior of the victim of the rape. Subjects who wrote mutations to the victim's behavior that would change the outcome assigned more fault (cause, blame, and responsibility) to her than subjects who wrote mutations that did not change the outcome. In another experiment, subjects listened to a mock lawyer's closing arguments suggesting possible changes to the story that would have changed the outcome. Again, if changing the victim's behavior would change the outcome, she was assigned more fault than if her actions would not change the outcome.

Alexandra Kincannon<sup>64</sup> has demonstrated a similar effect using materials that were loosely based on an actual event reported in the press in 1998.<sup>65</sup> All subjects read the following story:

Mr. Smith was driving drunk when his car struck Ms. Jones's car. Ms. Jones was seriously injured in the accident. On the way to the hospital, Ms. Jones refused to accept a blood transfusion in the ambulance, in accordance with her religious beliefs as a Jehovah's Witness. Ms. Jones was very insistent and told the EMTs repeatedly that she did not want a blood transfusion. At the hospital, the doctors informed Ms. Jones's family of the necessity of the blood transfusion and that her life was in danger, and

---

61. See Lawrence M. Solan & John M. Darley, *Causation, Contribution, and Legal Liability: An Empirical Study*, 64 LAW & CONTEMP. PROBS. 265 (Autumn 2001). See, e.g., *In re Agent Orange Prods. Liab. Litig.*, 597 F. Supp. 740 (E.D.N.Y. 1984).

62. See, e.g., Wells & Gavanski, *supra* note 35, at 161-62.

63. See Branscombe et al., *supra* note 7; see also Rachel McCloy & Ruth M.J. Byrne, *Thinking About What Might Have Been: If Only, Even If, Causality and Emotions*, in PROCEEDINGS OF THE TWENTY-FIRST ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY 373 (M. Hahn & S.C. Stoness eds., 1999); Michele R. Nario-Redmond & Nyla R. Branscombe, *It Could Have Been Better or It Might Have Been Worse: Implications for Blame Assignment in Rape Cases*, 18 BASIC & APPLIED SOC. PSYCHOL. 347 (1996); Richard L. Wiener et al., *Counterfactual Thinking in Mock Jury Assessments of Negligence: A Preliminary Investigation*, 12 BEHAV. SCI. & L. 89 (1994).

64. See Alexandra Kincannon, *Is Refusing Treatment the Same as Causing your Own Death?: How Counterfactual Alternatives Affect Causal Reasoning About Real-World Events* (1999) (unpublished study on file with author). The following materials are all from this study.

65. See Linda Deutsch, *Death of Witness is Manslaughter*, BOSTON GLOBE, Dec. 19, 1998, at A7.

they also refused the treatment. Ms. Jones died of her injuries a few hours later. Mr. Smith was charged with murdering Ms. Jones and his trial soon went to court.

After reading the story, subjects read a paragraph representing either the final argument of the prosecution or of the defense. In the prosecution's argument, the defendant's actions are portrayed as the single cause that made Ms. Jones's death unavoidable. In the defense's argument, several alternative counterfactual possibilities are suggested that would have prevented the death.

The prosecution argued:

Mr. Smith is directly responsible for the death of Ms. Jones. Although her religious beliefs prevented her from receiving a blood transfusion, she never would have been in such a situation if she had not been the victim of Mr. Smith's recklessness. Ms. Jones had no way of avoiding the accident. Mr. Smith chose to drive when he was drunk and his impairment caused him to strike Ms. Jones's car. As a result of that collision, Ms. Jones received her fatal injuries. Mr. Smith initiated the chain of events that led to her death and her religious convictions and medical treatment are irrelevant. Anyone could have pre-existing medical conditions or other complications that would interfere with typical treatment. Therefore, anyone could have been killed by Mr. Smith's actions and the only way death could have been prevented is if the accident had never occurred.

The defense argued:

Mr. Smith has acknowledged his responsibility for driving drunk and causing Ms. Jones's injuries. Although he caused the accident, he is not responsible for Ms. Jones's death. Her injuries were easily treatable with a blood transfusion, and if she had received the necessary transfusion, she would not have died. Ms. Jones's and her family's refusals to accept the transfusion were the ultimate cause of her death. Mr. Smith could not have known that Ms. Jones would not accept appropriate medical treatment and his action, although careless, was purely accidental. If his car had hit anyone other than Ms. Jones, or if Ms. Jones had decided to accept treatment, no one would have died as a result of this accident.

The forty-two subjects were then asked to write what they thought was the cause of the victim's death and to rate how much of a causal role each of the figures in the story played in that death. Subjects were asked to give their ratings using a scale from zero (not at all causal) to eight (the most important cause). Finally, subjects were asked whether they thought the accused drunk driver should receive a legal penalty and, if so, what it should be.

Causal ratings for the defendant (Mr. Smith) varied according to argument type. He was given a higher average causal rating in the pro-prosecution condition (mean = 6.2) than in the pro-defense condition (mean = 5.4), whereas causal ratings for the victim (Ms. Jones) were higher on average in the pro-defense condition (mean = 6.4) than in the pro-prosecution (mean = 5.0). The interaction between causal agent (Mr. Smith or Ms. Jones) and argument type was statistically significant.

#### B. Why Should Counterfactual Reasoning Affect Causal Reasoning?

From the above experiments, we believe that counterfactual reasoning affects causal reasoning. Why might this happen? Should this effect be relegated

to the realm of nonnormative heuristics in reasoning, or is there rational reason for it?<sup>66</sup>

We propose that counterfactual reasoning affects causal reasoning, because these two types of reasoning, although different, are related. Specifically, we believe that counterfactual reasoning provides important information for evaluations of causality, and the fact that counterfactual reasoning affects causal reasoning simply makes sense.<sup>67</sup>

Let's return for a moment to the multiple case scientist/probabilistic approach to studying causality. In that approach, researchers believe that people view things as causal to the extent that they raise the likelihood of the effect above some baseline. The baseline is observable because it has operated many times in the past. For example, we know how often flowers bloom without fertilizers being poured on them. In the single case "legal" approach, we argue that people view things as causal to the extent that they raise the likelihood of the effect above some baseline.<sup>68</sup> But what is the baseline when a set of events happens only once? In Spellman's experiments, subjects were provided with those probabilities;<sup>69</sup> where do subjects' probability estimates come from in the Branscombe et al.<sup>70</sup> and Kincannon<sup>71</sup> studies?

$$\begin{array}{l} \text{causal} \\ \text{contribution} \end{array} \rightarrow \begin{array}{l} \text{probability of the effect after} \\ \text{the cause occurs} \end{array} - \begin{array}{l} \text{probability of the effect before} \\ \text{the cause occurs} \\ \text{(baseline)} \end{array}$$

We believe that both pre-existing knowledge and counterfactual reasoning provide information to plug into the probability estimates in the single case approach. The above equation helps explain and predict several results in the counterfactual/causal literature.

First, why does considering alternatives that undo the outcome increase causal ratings? When subjects are asked to imagine ways in which an actor in a story might have done something differently so that the outcome would not

---

66. Much of the psychological literature on reasoning in the last thirty years is replete with examples of humans demonstrating nonnormative reasoning—that is, reasoning that varies from the normative rules of statistics, probability, and logic. For reviews, see JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman et al. eds., 1982); RICHARD E. NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT (1980). In the last 10 years, however, much of this so-called "irrational" behavior has been reinterpreted as "rational" or sensible, given the limitations of human processing abilities and the structure of information in the environment. See, e.g., JOHN R. ANDERSON, THE ADAPTIVE CHARACTER OF THOUGHT (1990); SIMPLE HEURISTICS THAT MAKE US SMART (Gerd Gigerenzer et al. eds., 1999); see also NEAL FEIGENSON, LEGAL BLAME: HOW JURORS THINK AND TALK ABOUT ACCIDENTS (2000) (reviewing some "irrational" heuristics and examples of how jurors use such heuristics in attributing blame).

67. See Spellman, *supra* note 6.

68. See Spellman, *supra* note 24.

69. See *id.*

70. See Branscombe et al., *supra* note 7.

71. See Kincannon, *supra* note 64.

have occurred, causality ratings for those actors *increase*.<sup>72</sup> Imagining alternatives that undo the outcome will create in subjects' minds more ways in which other things, besides the actual outcome, could have occurred. Thus, the imagined probability of the effect *before* the cause occurs is estimated as lower, so the causal contribution seems greater.

Second, why does considering alternatives that *do not* undo the outcome decrease causal ratings? When subjects are asked to imagine ways in which an actor in a story might have done something differently but the outcome would remain the same, causality ratings for those actors *decrease*.<sup>73</sup> Imagining alternatives that do not undo the outcome will make the imagined probability of the effect *before* the cause occurs be estimated as higher, so the causal contribution seems lower.

Third, the Jehovah's Witness experiment provides a slight variation on changing subjects' probability estimates. What the defense argument did was change the subjects' estimates of the probability of the effect after the cause occurs. In the study, subjects may have some pre-existing belief as to how likely it is that someone would die as a result of the kind of car accident described. As they imagine that a transfusion, a normal occurrence, would have saved the victim, their estimates as to how likely death was *after* the accident decreases. Therefore, the perceived causal contribution of the driver decreases.

Fourth, the equation makes some other predictions regarding the relation between mutability and causality. For example, the equation predicts that the more alternative choices one has, some of which would lead to a different outcome, the more causality will be attributed to an event that causes the actual outcome to occur. Such a result is illustrated in the Wells and Gavanski wine experiment.<sup>74</sup> When the boss considered choosing between two dishes, only one of which contained wine, his choice was rated as more causal of the employee's death than when he considered choosing between two dishes that both contained wine. Increasing the number of alternatives that would not lead to the outcome would decrease the baseline and thus should increase causality. The equation above predicts this result not only for experiments in which there is a 50% chance (one-wine condition) versus a definite 100% chance (two-wine condition) of choosing a bad option, but also for conditions with varying numbers of choices. Recently, Barbara Spellman and Nadine M. Meyers have found support for this prediction.<sup>75</sup> In one experiment, subjects read about a restaurant in which either one-of-five, three-of-five, or all five-of-five dishes that one person considered ordering for another contained a "bad" ingredient. Such a dish is ordered and the nonordering person gets food poisoning. The person ordering is considered most causal in the condition in which only one dish had

---

72. See Branscombe et al., *supra* note 7; McCloy & Byrne, *supra* note 63.

73. See Branscombe et al., *supra* note 7; McCloy & Byrne, *supra* note 63.

74. See Wells & Gavanski, *supra* note 35, at 163-65.

75. See Barbara A. Spellman & Nadine M. Meyers, Wine, Women, and Food Poisoning: The More-Choices Leads to More-Causality Effect in the Relation Between Counterfactual and Causal Reasoning (2000) (unpublished manuscript, on file with author).

the ingredient; less so when three dishes had the ingredient; and even less than that when all dishes had the ingredient.

## V

### IMPLICATIONS FOR JURORS' DECISIONS

The relation between counterfactual and causal reasoning has implications for at least two potential components of a jury trial: the presentation of counterfactual counter-arguments during trial and instructions on causality.

#### A. Arguments in the Courtroom

The research that shows that considering counterfactual alternatives to an outcome may affect causal reasoning has clear implications for the courtroom.<sup>76</sup> Using counterfactuals to show that an outcome was, *a priori*, very unlikely to happen, will make the actors who brought about that outcome seem more causal. Using counterfactuals to show that an actor could have made other choices that would have resulted in an alternative outcome will also make that actor seem more causal.

There are certainly limits as to the range of usefulness of this technique. First, it does not always work. Our laboratory has data from an experiment in which suggesting counterfactual alternatives to an action that would undo an outcome did *not* increase liability judgments against a defendant; however, in that experiment, there was no reason to find the defendant liable at all.<sup>77</sup> Second, even if the technique does make an actor seem more causal, that fact does not imply that he would be more likely to be judged a cause in court. As noted by Robinson and Darley, although people treat causality as continuous, the law treats it as dichotomous.<sup>78</sup> Third, in a courtroom, when each side can present counterfactual arguments of the “only if” and “even if” varieties, the effects of hearing conflicting counterfactuals might cancel out (or produce some other unpredictable result). Thus, exactly how effective this technique is, and how to counter it, are certainly areas for more research.

#### B. Jury Instructions

Given what we know from the experimental literature about how people follow jury instructions, what should we expect from instructions on causality?

---

76. See FEIGENSON, *supra* note 66, at 125, 155 (providing examples of lawyers' actual use of counterfactual arguments).

77. Subjects read about a woman whose car was stopped at a red light. She fiddled with the radio station and therefore hesitated before starting when the light turned green. A line of cars with a school bus at the end then accelerated behind her. When the bus got to the intersection, a car driven by an upset man who had been fired that day came through the light in the other direction and hit the bus, injuring many children. Some subjects read about a lawsuit against the woman. Even when they were reminded that had she not hesitated, the accident would not have occurred (that is, mutating her actions would change the outcome), no subjects thought she should be held liable or should pay any damages. See Barbara Spellman, *Untitled Research* (1997) (unpublished data on file with author).

78. See ROBINSON & DARLEY, *supra* note 58, at 208-10.

Our expectations are derived from the studies described above and three relevant studies involving jury instructions: In one, no instructions on causality were given; in another, instructions on causality seemed to have no effect; and in the third, instructions on the definitions of crimes themselves seemed to have no effect.

First, when subjects acting as mock jurors making decisions about punitive damages were not instructed at all on issues of causality, some subjects (60%) spontaneously raised issues of causality, and of those, many (58%) wrote as if they were considering “but for” causality to be their guide.<sup>79</sup> On the other hand, the double murder and storm experiments, described above, suggest that subjects do not always spontaneously use “but for” causal reasoning—at least when it is not appropriate in cases of multiple sufficient causes.

Second, in a different set of studies, subjects read one of two versions of scenarios involving various legal controversies. The versions differed on causal proximity; that is, how long the causal chain was from the initial (negligent) action to the outcome. For example, one set of materials was described as follows:

In the handgun case a businessman left a briefcase containing handguns in an empty chair of an airport waiting area. In the simple-causal-chain version, the briefcase was stolen by a maladjusted individual who used one of the guns to shoot and severely injure a taxi driver. In the complex-causal-chain version, the briefcase was taken by a security guard to the lost-and-found department, where the gun was stolen by the lost-and-found clerk. The maladjusted individual later picked this gun from the pocket of the clerk and used it to shoot the taxi driver. In both versions the driver sued the businessman for his injuries.<sup>80</sup>

Subjects were given a set of jury instructions to read and were told to be guided by them. The instructions were the California instructions on negligence (BAJI 3.00, 3.10, 3.11, and 3.75) which included definitions of negligence and proximate cause. The instructions did not mention how jurors should consider the length of a causal chain. However, in the second experiment the subjects also read that “a person . . . is not relieved of liability because of the intervening act of a third person (or persons) if such act was reasonably foreseeable at the time of his negligent conduct.”<sup>81</sup>

In both experiments, subjects rated the injury as being less foreseeable and the defendant as being less liable in the scenario with the longer causal chain.<sup>82</sup> The authors suggest that “one factor contributing to the causal proximity effect may have been some subjects’ disregard of these instructions.”<sup>83</sup> Of course, one

---

79. See Reid Hastie, *The Role of “Stories” in Civil Jury Judgments*, 32 U. MICH. J. L. REFORM 227, 236 (1999). Note that subjects in this experiment were not students, but rather “citizens from the Denver area.”

80. Joel T. Johnson & Jerome Drobny, *Proximity Biases in the Attribution of Civil Liability*, 48 J. PERSONALITY & SOC. PSYCHOL. 283, 285 (1985).

81. *Id.* at 288.

82. See ROBINSON & DARLEY, *supra* note 58, at 181-89 (relaying a study in which distance matters, but not systematically).

83. Johnson & Drobny, *supra* note 80, at 293.

problem with this experiment for us is that there was no condition in which subjects were *not* given instructions; therefore, we cannot actually evaluate the effect of the instructions per se.

A third very interesting set of studies by Vicki L. Smith suggests that, when subjects act as mock jurors when evaluating criminal cases, they rely more on their commonsense or “prototypical” views of what constitutes a particular crime than on the judge’s instructions regarding the elements of the crime.<sup>84</sup> Smith had some subjects describe what they thought were the elements of various crimes including assault, burglary, kidnapping, murder, and robbery. She found that people had ideas of what these crimes would generally be like (that is, “prototypes”), and that when subjects read scenarios about crimes that were closer to the prototype, they were more likely to convict the perpetrator.<sup>85</sup> More relevant, she found that subjects who were given instructions on the elements of the crime before reading the scenario made the same judgments as subjects who did not receive instructions about the elements of the crime. These subjects relied on their pre-existing knowledge or beliefs about what constituted a particular crime, rather than the judge’s instructions on what constituted the crime.<sup>86</sup>

As with the more common classes of crime, people have pre-existing knowledge and beliefs about causality. We believe that researchers should try to understand those beliefs and learn how people reason naturally when making causality judgments in the absence of jury instructions. The legal system, however, will no doubt insist upon instructing juries about causality (whether or not it finds such uninstructed judgments lacking). What is likely to happen when those pre-existing beliefs collide with jury instructions? On one hand, it seems that we should infer from Smith’s studies that, in any clash, people will rely on their own beliefs rather than on instructions when evaluating causality. On the other hand, this legal misreliance does not sound as inappropriate as it does for evaluating the elements of a crime. Like many of the issues addressed by Robinson and Darley,<sup>87</sup> we believe that causality might be an issue where community views and people’s moral intuitions *ought* to be guiding legal standards. In fact, the countless instructions and rulings regarding issues of causation (especially with the notion of “proximate cause”) seem to be an attempt to capture exactly those societal views of fairness.

“Proximate cause” . . . is merely the limitation which the courts have placed upon the actor’s responsibility for the consequences of his conduct. In a philosophical sense, the consequences of an act go forward to eternity, and the causes of an event go back to the discovery of America and beyond. . . . As a practical matter . . . , some boundary

---

84. See Vicki L. Smith, *Prototypes in the Courtroom: Lay Representations of Legal Concepts*, 61 J. PERSONALITY & SOC. PSYCHOL. 857 (1991) [hereinafter Smith, *Prototypes in the Courtroom*]; Vicki L. Smith, *When Prior Knowledge and Law Collide: Helping Jurors Use the Law*, 17 LAW & HUM. BEHAV. 507 (1993).

85. See Smith, *Prototypes in the Courtroom*, *supra* note 84, at 868-70.

86. See *id.*

87. ROBINSON & DARLEY, *supra* note 58.

must be set to liability for the consequences of any act, upon the basis of some social idea of justice or policy.<sup>88</sup>

This limitation is sometimes, although rather infrequently, one of the facts of causation. More often it is purely one of policy, of our inadequately expressed ideas of what justice demands, or of administrative efficiency and convenience—none of which have any connection with questions of causation at all.

## VI

### CONCLUSION

Counterfactual reasoning is essential for causal reasoning in the law. The psychological literature in the past assumed that counterfactual reasoning was, in fact, the basis of all causal reasoning; however, recent research shows that lay causal reasoning is not that simplistic. In particular, in experiments involving complicated cases of multiple sufficient causes, we have seen that subjects do distinguish between “but for” causes (or what they choose to mutate) and actual causes. In fact, subjects come up with the legally correct judgment in the absence of instructions on causality.

There is also psychological evidence demonstrating that counterfactual reasoning may affect causal reasoning in legally relevant situations. Listening to others generate counterfactual alternatives to an outcome, or generating such alternatives oneself, can affect one’s causal judgments. Rather than considering this effect to be irrational, we have argued that it is quite systematic and have explained a possible rational basis for it. When considering causal judgments for the kinds of simple scenarios used in these “single case ‘legal’ approach” studies, we have great respect for uninstructed lay responses.

Amusingly, the fact that uninstructed people handle the “causal conundrum” of multiple sufficient causes quite sensibly, despite the huge number of pages written about it, would not be surprising to many of the legal commentators who have written so many of those pages. Hart and Honoré state that, “causal judgments, though the law may have to systematize them, are not specifically legal. They appeal to a notion which is part of everyday life and which ordinary people, including, jurymen, can handle with a minimum of guidance.”<sup>89</sup> Regarding “causation in fact,” in particular, Prosser & Keeton conclude that “[i]t is a matter upon which lay opinion is quite as competent as that of the most experienced court.”<sup>90</sup>

---

88. PROSSER, *supra* note 48, at 236-37.

89. HART & HONORÉ, *supra* note 32, at lv.

90. PROSSER & KEETON, *supra* note 49, at 264.