

Judgment Dissociation Theory: An Analysis of Differences in Causal, Counterfactual, and Covariational Reasoning

David R. Mandel
University of Victoria

Research suggests that causal judgment is influenced primarily by counterfactual or covariational reasoning. In contrast, the author of this article develops judgment dissociation theory (JDT), which predicts that these types of reasoning differ in function and can lead to divergent judgments. The *actuality principle* proposes that causal selections focus on antecedents that are sufficient to generate the actual outcome. The *substitution principle* proposes that ad hoc categorization plays a key role in counterfactual and covariational reasoning such that counterfactual selections focus on antecedents that would have been sufficient to prevent the outcome or something like it and covariational selections focus on antecedents that yield the largest increase in the probability of the outcome or something like it. The findings of 4 experiments support JDT but not the competing counterfactual and covariational accounts.

If causation is the cement of the universe, as the philosopher David Hume (1740/1938) put it, then it is fair to say that causal knowledge is the cement that binds together each person's representational universe. Causal reasoning—the process that generates this glue—confers many functional advantages. In virtually every sphere of human interest, our abilities to learn and categorize (Rehder & Hastie, 2001); to explain the past, predict the future, and exert control in the present (Weiner, 1985); and to construct meaning and make decisions by synthesizing coherent stories (Pennington & Hastie, 1992; Read, 1987) depend on the ways in which we reason about the causes and effects of events around us. Hence, it should come as no surprise that the study of causal reasoning has been of great interest to psychologists in a broad spectrum of fields, such as animal learning (Pearce & Bouton, 2001), perception (Michotte, 1946/1963), cognition (Keil & Wilson, 2000; Sperber, Premack, & Premack, 1995), social psychology (Heider, 1958; Kelley, 1973), developmental psychology (Bullock, Gelman, & Baillargeon, 1982; Piaget, 1974), and methodology (Cook & Campbell, 1979).

Causal reasoning takes multiple forms. First, causal reasoning can focus on inducing causal regularities over a set of cases or on

judging the probable cause of an outcome in a particular case (e.g., see Cheng, 1993; Hart & Honoré, 1985; Hilton, 1990; Kelley, 1973; Spellman & Mandel, 2003). The focus of this article is on causal judgments made about specific cases. The present analysis would thus apply to reasoning about the causes of, say, a particular fire, but not to reasoning about the causes of fire in general. This type of case-based causal reasoning is pervasive in everyday social perception (Kelley, 1973) as well as in law (Greene & Darley, 1998; Spellman & Kincannon, 2001) and political science (Tetlock & Belkin, 1996). Second, causal reasoning can focus on understanding the causal factors in a particular case or on selecting from those factors what is to be deemed the cause. Unlike causal understanding, causal selection is conversational in nature and often involves attempting to bridge a gap in an explainee's knowledge (Hilton, 1990; Hilton & Erb, 1996; McGill, 1989; McGill & Tenbrunsel, 2000). For example, if asked why a nearby restaurant burned down, you might mention a rumor that you heard about the restaurant owners intentionally setting fire to the restaurant because they were having financial difficulties rather than the fact that oxygen was a necessary contributing factor. Given that oxygen is always present when fires occur, mentioning this factor does not add to what the explainee likely knows. The focus of this article is on the type of factors that explainers select when asked causal and related types of questions.

In the present article, I examine two contemporary accounts of causal selection and then propose a new account. Briefly, the *counterfactual simulation account* (CSA) posits that reasoners select the cause of a particular outcome by running a counterfactual simulation in which the proposed cause is mentally negated and the effect on the outcome is then assessed (e.g., Kahneman & Tversky, 1982a). If the outcome is also mentally negated—or “undone”—then the proposed cause is more likely to be selected as the cause. Otherwise, the proposed cause will not be selected and the search for an adequate explanation will continue. The second account, proposed by Spellman (1997), posits that causal selection relies primarily on a form of covariational analysis, akin to a multiple regression analysis in which the predictors (i.e., proposed causes) are entered into the model in chronological order (Hilton,

This research was funded in part by a University of Victoria research grant and Research Grant 249537–02 from the Natural Sciences and Engineering Council of Canada. Part of this research was presented on May 16, 2001, at the European Association of Experimental Social Psychology Small Group Meeting on Counterfactual Thinking, Aix-en-Provence, France, and on July 28, 2002, at the 13th General Meeting of the European Association of Experimental Social Psychology, San Sebastian, Spain.

I thank Zoë Crossley, Farish Sawyer, Penny Silsbury, Karen Souza, Rebecca Turner, and Clare Marsh for their assistance with this research. I also am grateful to Mandeep Dhimi and Adam Galinsky for their valuable feedback on earlier versions of this article and to Michael Hunter for statistical advice.

Correspondence concerning this article should be addressed to David R. Mandel, Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, British Columbia V8W 3P5, Canada. E-mail: dmandel@uvic.ca

1988). According to Spellman's probability-updating account (SPA), once the outcome of a case is known, causal reasoners retrace the path of critical events that led up to it and intuitively assess how each one uniquely raised the outcome's probability. The event that is perceived to raise the outcome's probability the most will be selected (or "credited," in Spellman's terms) as the cause.

The new account, which I refer to as *judgment dissociation theory* (JDT), is strongly motivated by its predecessors. It focuses not only on explaining how the causal selection process works but also on explaining how it differs from the selection process in counterfactual reasoning and the type of covariational judgment process proposed by SPA. The theory assumes that counterfactual and covariational reasoning can influence the causal selection process. However, it also proposes that causal, counterfactual, and covariational reasoning differ in important respects that will lead to predictable dissociations in the focus of the three corresponding types of judgment under certain boundary conditions. In the present research, I examine cases that conform to a predicted boundary condition for a three-way judgment dissociation in order to test the limits of CSA and SPA and to examine whether JDT could account for deviations from the predictions of these accounts. In the next two sections, I examine CSA and SPA in greater detail, after which I introduce the core ideas behind JDT and examine how its predictions differ from CSA and SPA. Four experiments that pit the three accounts against each other are then reported.

CSA

Like causal reasoning, counterfactual thinking is ubiquitous in everyday life (Hofstadter, 1979). People often think about what almost happened (Kahneman & Varey, 1990) and what "coulda, shoulda, or woulda" happened if only something had happened differently (Kahneman & Tversky, 1982a). Over the past two decades, a great deal of psychological research has examined the determinants, consequences, and functions of counterfactual thinking (for a review, see Roese, 1997). In particular, attention has focused on counterfactual conditionals that take the form of "if-only" thoughts (e.g., "if only I had taken my usual route home today, I wouldn't have been in this car accident"). It has been proposed that these counterfactuals (often called *upward* counterfactuals because they bring to mind better possible worlds) play an important role in planning for the future by spotlighting causal contingencies (Markman, Gavanski, Sherman, & McMullen, 1993; Roese, 1994).

According to CSA (e.g., Kahneman & Tversky, 1982a; Lipe, 1991; Roese & Olson, 1997; see also Mackie, 1974), people often infer the causes of past outcomes by running mental simulations in which the proposed cause is negated and the outcome in the simulation is observed. If the outcome is mentally undone as a result of mutating the proposed cause, then the probability that that factor will be selected as the cause should increase. Alternatively, if the outcome is not undone, then the proposed cause is unlikely to be selected. Support for CSA was provided by Wells and Gavanski (1989; see also Branscombe, Owen, Garstka, & Coleman, 1996). In their second experiment, participants read about a cab driver who refused the fare of a paraplegic couple. The couple decided to drive their own specially equipped car and were

killed when the bridge that they drove over collapsed. In the low-mutability version of the story, the cab driver also drove off the bridge, whereas he made it across safely 15 min before the couple reached the bridge in the high-mutability version. Participants listed four ways that the couple's death could have been avoided and rated the causal importance of the cab driver's refusal to take the couple. Demonstrating the importance of perceived mutability in both counterfactual and causal reasoning, participants in the high-mutability condition listed the cab driver's refusal more often as a way of undoing the outcome and ascribed greater causal importance to it than did participants in the low-mutability condition (cf. Mandel & Lehman, 1996, Experiment 3). Moreover, consistent with the idea that counterfactual simulations influence causal judgments, participants who listed the cab driver's refusal to take the couple in their counterfactuals ascribed greater causal importance to the refusal than did participants who did not list this factor.

However, several studies (e.g., Davis, Lehman, Wortman, Silver, & Thompson, 1995; Mandel & Lehman, 1996; for a review, see Spellman & Mandel, 1999) have called CSA into question. For example, N'gbala and Branscombe (1995, Experiment 2) created a more complex version of the cab scenario that included two cab drivers and a controllability variable. The first cab driver does not take the couple because he made a mistake and picked up the wrong couple (controllable) or because his car broke down (uncontrollable). The couple then took another cab that crashed off the bridge either because the second cab driver was drunk (controllable) or because the bridge collapsed (uncontrollable). Participants generated a single counterfactual (i.e., the outcome could have been different if only. . .) and rated the extent to which the first cab driver, the second cab driver, the bridge, and the couple were responsible, blameworthy, and causally implicated. Although more participants listed the first cab driver than either the second cab driver or the collapsed bridge in their counterfactual selections, greater fault (a composite of the ratings) was assigned to the second cab driver (when he was drunk) or to the collapsed bridge. Thus, when participants are given the opportunity to focus on different factors in their counterfactual and causal judgments, they sometimes do (see also Mandel & Lehman, 1996). Moreover, contrary to CSA, participants who mutated (i.e., mentally changed) a particular antecedent did not rate it as more at fault than did participants who did not mutate it. More recently, Mandel (2003b) directly compared the effects of counterfactual thinking and factual thinking on causal judgment. Using both recollected real-life and hypothetical events, Mandel found that participants directed to think counterfactually about an actor's role in the relevant case did not assign greater causal importance to that actor than did participants who were directed to think factually about the actor.

Furthermore, a significant limitation of CSA is its implication that causal reasoners focus primarily on necessary causation. If A and B represent a proposed cause and a target outcome, respectively, and if A is a necessary facilitative cause of B, then the negation of A implies the negation of B ($\neg A \supset \neg B$; Cummins, 1995; Fairley, Manktelow, & Over, 1999). This conditional relation is akin to the counterfactual conditional in which the mental negation of A undoes B. Necessary conditions for an outcome, such as the presence of oxygen for fire, are more likely to be viewed as enabling conditions than causes (e.g., Cheng & Novick, 1991; Hilton, 1988). Instead, reasoners are more likely to focus on

causes that are perceived to be sufficient under the circumstances. Two lines of evidence support this assertion. First, studies of naive causal understanding (e.g., Goldvarg & Johnson-Laird, 2001) indicate that people define causation primarily in terms of sufficiency. Mandel and Lehman (1998, Experiment 1) found that the vast majority of participants asked to define the terms *cause* and *preventor* in their own words provided a sufficiency definition but not a necessity definition for each term. For example, a common definition of *cause* was “if the cause happens, then the effect will happen too” (i.e., sufficiency). Few participants stated anything like “if the cause doesn’t happen, then the effect won’t either” (i.e., necessity). Second, studies that examine how people integrate contingency information when asked to judge the magnitude of a causal relation show that people assign greater weight to cases in which the proposed cause is present rather than absent (Anderson & Sheu, 1995; Cheng, 1997; Kao & Wasserman, 1993; Mandel & Lehman, 1998; McGill, 1998; McGill & Klein, 1993). The cause-present information is required to compute the probability of the outcome given the proposed cause, $P(B|A)$, which is regarded as a probabilistic measure of the sufficiency relation $A \supset B$ (e.g., Hilton, 1988; Suppes, 1970). In contrast, the cause-absent information is required to compute the probability of the outcome given the absence of the proposed cause, $P(B|\neg A)$, which is regarded as a probabilistic measure of necessity.

In summary, the extant research does not provide much support for CSA. However, an important test of CSA remains to be conducted. If the relation between causal and counterfactual selections is driven by a “counterfactual-influences-cause” process, then two asymmetric effects should be observed. First, counterfactual judgments should facilitate causal judgments more than vice versa. Roese and Olson (1997) reported an experiment that demonstrated this pattern. However, an unpublished experiment by Mandel, N’gbala, and Bonnefon (2001) did not find that counterfactual selections facilitated causal selections. Second, the correlation between importance ratings attached to causal and counterfactual selections should be greater when the counterfactual task precedes the causal task than vice versa. One objective of Experiments 1 and 2 in the present research was to conduct the first systematic test of the latter prediction.

SPA

Several theorists have proposed that covariational assessments provide an important, if not primary, cue to causation. Most of these accounts are designed to describe causal induction over a set of cases considered by a reasoner in which the cause and the effect may be present or absent (e.g., Anderson & Sheu, 1995; Cheng, 1997; Kelley, 1967; Schustack & Sternberg, 1981). In contrast, SPA was designed to predict causal selections within a single case (for related accounts, see Brewer, 1977; Hilton, 1988; Hogarth & Einhorn, 1992). As noted above, SPA proposes that causal reasoners select the cause of an outcome by performing a series of intuitive probability assessments that are designed to identify the proposed cause that raises the probability of the outcome by the greatest factor. To visualize how SPA works, consider the coin-toss problem (Miller & Gunasegaram, 1990): Two players can win a cash prize if they each toss a coin that lands on the same side; otherwise, they win nothing. Player 1’s coin lands on heads, then Player 2’s coin lands on tails. So they win nothing. According to

SPA, one would first estimate the outcome’s prior probability, which is $P(\$) = .5$. The second step is to estimate the change in the outcome’s probability given the occurrence of the first proposed cause. Let ΔP_i refer to the change in probability due to a particular proposed cause, where i enumerates the temporal order of the proposed causes in a hypothesized set. In the present example, this set includes two proposed causes, Player 1 and Player 2. Given that $P(\$|Player\ 1\ tosses\ heads) = .5$, $\Delta P_1 = 0$. However, given that $P(\$|Player\ 1\ tosses\ heads \wedge Player\ 2\ tosses\ tails) = 1$, $\Delta P_2 = .5$ (\wedge expresses an event conjunction and is read as “and”). According to SPA, then, Player 2 will be deemed the cause. In fact, most participants do judge that Player 1 will blame Player 2 more than vice versa for the unappealing outcome (Byrne, Segura, Culhane, Tasso, & Berrocal, 2000; Miller & Gunasegaram, 1990; Spellman, 1997).¹

In a more direct test of SPA, Spellman (1997, Experiment 2) asked participants to consider a game show in which two team players have the opportunity to win a \$10,000 prize. Player 1 has to fill a balloon with water in less than 10 s and Player 2’s task depends on whether Player 1 was successful. Spellman manipulated the degree to which Player 1 and Player 2 changed the probability of the outcome of the game (either winning or losing the prize) so that in some conditions Player 1 contributed the most (i.e., $\Delta P_1 > \Delta P_2$) and in other conditions Player 2 contributed the most (i.e., $\Delta P_1 < \Delta P_2$). Participants were then asked to rate the contribution of each player to the final outcome on separate 0–10 scales. With one exception, contribution ratings increased monotonically with the change in outcome probability due to the rated target. However, Spellman also obtained results that clearly indicate the incompleteness of SPA. First, in one set of conditions, ΔP_1 was held constant at 0 while ΔP_2 was varied. Spellman compared conditions in which ΔP_2 was equal to either .1 or .9 and found that in spite of the extreme difference in Player 2’s contribution in these two sets of conditions, ratings of Player 2’s contribution differed only marginally between them. Second, Spellman was able to compare one set of conditions in which $\Delta P_1 > \Delta P_2$ with another set of conditions in which $\Delta P_1 < \Delta P_2$ by the same value. For example, in one set of conditions, $\Delta P_1 = .1$ and $\Delta P_2 = .4$, whereas in another set of conditions, $\Delta P_1 = .4$ and $\Delta P_2 = .1$. Her data clearly reveal that after controlling for ΔP , Player 2 was rated as having a greater contribution than Player 1. This effect of temporal order indicates that proposed causes that are proximal to the outcome are judged to be of greater causal importance than are distal causes for reasons other than those having to do with changes in outcome probability. I shall return to this finding in the next section, where I develop JDT.

Although SPA posits that causal selection depends on subjective probability assessments, no study has examined the relation between subjective probability and causal judgment to ascertain whether SPA’s predictions are met. Therefore, a second objective of the present research was to conduct this critical test. Furthermore, the present research examined an intriguing prediction of SPA that, until now, has also not received attention. That is, SPA implies an asymmetric form of causal discounting that favors the selection of proposed causes that occur earlier in a case even when

¹ SPA assumes that subjective rather than objective probabilities would provide the input to the causal selection process.

they are matched by later causes in terms of their sufficiency for producing the outcome (cf. Ahn & Bailenson, 1996; Kelley, 1973). In cases involving multiple sufficient conditions for an outcome, SPA predicts that the earliest sufficient condition will invariably be selected as the cause. For example, imagine a case in which a man unknowingly ingests a slow-acting but lethal dose of poison. Before the poison has a noticeable effect, the man is killed by a gunshot wound. According to SPA, the poisoning would be selected as the cause of the man's death because it would lead to a dramatic increase in outcome probability (i.e., ΔP_1 would be very high). This would create a ceiling effect whereby there would be little room for that probability to increase further by updating it with the gunshot (i.e., ΔP_2 would have to be very low). This prediction is at odds with intuition, which suggests that most people would select the gunshot rather than the poisoning as the cause of death. I tested this prediction of SPA and pitted it against JDT in Experiments 1 and 2 of the present research.

JDT

Consistent with past functionalist accounts (Roese, 1994; Taylor, 1991; Weiner, 1985), JDT posits that causal, counterfactual, and covariational reasoning are adaptive cognitive processes that can aid in prediction, control, and explanation across many contexts. Nevertheless, JDT draws functional distinctions between these types of reasoning and predicts that under some conditions, there will be clear dissociations in the foci of causal, counterfactual, and covariational judgments. In contrast to CSA but consistent with the evidence reviewed above, JDT posits that selected causes are likely to be perceived as sufficient under the circumstances. According to JDT, causal selection is also particularly likely to focus on events that played a decisive role in how the outcome of a case was actually generated. In contrast to SPA, the theory posits that causal selections will be perceived as informative to the extent that they elucidate the causal mechanism that generated the outcome. Hence, JDT draws conceptual connections to accounts that posit the importance of perceived mechanisms (Ahn & Kalish, 2000; Ahn, Kalish, Medin, & Gelman, 1995; Harré & Madden, 1975; Michotte, 1946/1963; Shultz, Fisher, Pratt, & Rulf, 1986) and perceived changes of propensity (Kahneman & Varey, 1990) in causal reasoning. This may be why Spellman's (1997, Experiment 2) participants ascribed greater importance to the proximal cause even after controlling for probability. Whereas Player 1 merely determined what game Player 2 would play, Player 2 directly influenced the actual outcome.

Unlike CSA, JDT proposes that the explanatory focus of counterfactual reasoning is more likely to be placed on prevention than on facilitative causation. This view is consistent with research indicating that upward counterfactuals are likely to be recruited in response to negative outcomes (Roese, 1997; Roese & Olson, 1997). By learning how a negative outcome might have been prevented, one may also plan to improve achievement in the future (Markman et al., 1993; Roese, 1994). Mandel and Lehman (1996) found direct support for the idea that counterfactuals focus on prevention. In their first experiment, participants read a story about a man (Mr. Jones) who decided to drive home one day by a different route. Along the way, a drunk driver who failed to stop at a red light rammed into Mr. Jones's car, seriously injuring Mr. Jones. One group listed ways that Mr. Jones would probably

complete the thought "If only. . . ." A second group was asked to list what Mr. Jones probably thought about when he contemplated how the accident could have been prevented. And a third group was asked to list what Mr. Jones probably thought about when he contemplated how the accident was caused. Participants in the counterfactual and prevention conditions focused primarily on Mr. Jones's decision to take the unusual route home. In contrast, participants in the causation condition were significantly more likely to focus on the drunk driver's actions than were participants in the other conditions. The dissociation between causal and counterfactual selections is also revealed by the attributional analyses that unfolded in the months after the 9/11 terrorist attacks. Although the cause of the attacks was attributed to the terrorists who carried them out and to al Qaeda, most counterfactual assessments have been directed at prevention failures by U.S. intelligence agencies.² Finally, note that by shifting the proposed emphasis of counterfactual reasoning to the selection of preventors rather than causes, JDT reconciles the view that counterfactuals do tell us something of a causal nature (broadly speaking), with the evidence reviewed above indicating that causal judgment focuses more on sufficiency than necessity. That is, JDT posits that the negated antecedent in if-only reasoning represents the affirmation of a sufficient but forgone preventor rather than the negation of a necessary cause that did, in fact, occur in the relevant case.

A key proposal of JDT is that causal, counterfactual, and covariational judgments differ with respect to their specification of the outcome of a particular case. As noted above, JDT proposes that causal selection focuses on explaining the actual outcome of a case rather than a similar type of outcome that may also have been possible or even probable. I refer to this idea as the *actuality principle* in causal reasoning. In contrast, JDT also proposes that counterfactual and covariational judgments focus on an ad hoc category of outcomes in which the actual outcome represents the category prototype or norm. I refer to this idea as the *substitution principle* in counterfactual and covariational reasoning. Ad hoc categories (e.g., *things to bring to today's meeting*) are constructed in response to specific, short-term goals that, once satisfied, do not require the category to be maintained (Barsalou, 1983, 1991). In JDT, the goals that give rise to ad hoc categories of outcomes are associated in predictable ways with different reasoning processes. In counterfactual reasoning, the category exemplars must satisfy the goal of finding ways to undo X or something like it. In covariational reasoning, the category exemplars must satisfy the goal of finding ways to increase the likelihood of X or something like it. More generally, the substitution principle coheres with Kahneman and Tversky's (1982b) idea that "it is frequently appropriate in conversation to extend the definition of an event X to 'X or something like it'" (p. 149). The substitution principle generalizes this idea by proposing that a similar extension from X to X or something like it characterizes the mental representations that people use in counterfactual and covariational reasoning.

² For instance, one *New York Times* article began by stating, "The director of the F.B.I., Robert S. Mueller III, acknowledged today for the first time that the attacks of Sept. 11 might have been preventable if officials in his agency had responded differently to all the pieces of information that were available" (Lewis, 2002, ¶ 1).

Example of JDT's Predictions

To visualize how causal, counterfactual, and covariational judgments may be dissociated, consider a case in which a professional criminal falls prey to two assassination attempts. Assassin 1 slips the victim a lethal dose of poison that will kill him in 1 hr but that will have no effect for the first 30 min. Before the poison takes effect, Assassin 2 runs the victim off the road. The car explodes and the victim is killed. SPA predicts that Assassin 1 would be selected as the primary causal agent of the victim's death because the subjective probability of the victim's death should rise to near certainty after the first assassination attempt. In fact, SPA predicts that people would ascribe little causal importance to Assassin 2 because there could be relatively little increase in the probability of the victim's death after controlling for the poisoning. In contrast, JDT predicts that even though the lethal poison was consumed prior to the car crash and was sufficient to kill the victim, most people will judge the subsequent car crash to be of greater causal importance because only the second assassination attempt played a direct role in how the victim actually died. Unlike SPA, therefore, JDT does not assume that discounting favors earlier causes over later ones even when they are of comparable sufficiency.

The predictions of JDT also differ from several predictions derived from the literature on counterfactual thinking. First, CSA predicts that the rank order of causal importance judgments for the relevant proposed causes will follow the same ranking as for counterfactuals, especially if counterfactual assessments are made right before causal assessments. Second, Wells, Taylor, and Turtle (1987) proposed that people will tend to attribute causation to initial events in a chain of events leading up to a focal outcome because earlier events in the chain are more mutable given that they are not constrained by later events (see also Vinokur & Ajzen, 1982). Third, and in opposition to the preceding "primacy-constraint hypothesis," Miller and Gunasegaram (1990; Kahneman & Miller, 1986) proposed that people will tend to attribute causation to the most recent event in a chain because it will be the most mutable. To reconcile their prediction with that of Wells et al. (1987), they proposed that their "recency hypothesis" applies only to chains of events in which there are no obvious causal links between the antecedents themselves. Both the primacy-constraint and recency hypotheses predict that counterfactual and causal listings or importance ratings will correspond in rank order of frequency or strength. In contrast, JDT predicts that causal and counterfactual judgments are likely to be dissociated; for example, in the scenario given above, the mental negation of the car crash is likely perceived as insufficient to prevent the victim's death. If the substitution principle applies to counterfactual reasoning, then the most parsimonious way to undo the victim's death would be by mutating the victim's involvement in crime, which set the stage for both assassination attempts. The objective of Experiments 1 and 2 was to test the key predictions of JDT by pitting them against those of SPA, CSA, and related hypotheses.

Experiment 1

In Experiment 1, I examined the relations between causal, counterfactual, and covariational judgments about the assassination case just described. In line with JDT, I predicted a three-way judgment dissociation, such that (a) causal judgments would focus

primarily on the car crash, which played a direct role in generating the actual outcome; (b) counterfactual judgments would focus primarily on negating the victim's involvement in crime, a mutation that is sufficient to (mentally) prevent both the inevitable death by poisoning and the actual death by car crash; and (c) the greatest change in outcome probability would be associated with the lethal poison, which represented the first sufficient condition for the victim's death. Experiment 1 also provided the first test of SPA that relied on subjective probability estimates. For brevity, I refer to (a) the victim's involvement in organized crime, (b) the first assassination attempt (poisoning), (c) the second assassination attempt (car crash), and (d) the outcome (the victim's death) as CRIMELIFE, POISON, CRASH, and DEATH, respectively.

Method

Participants. One hundred fifty-three undergraduate students participated in the experiment on a voluntary basis.

Materials and procedure. Participants were presented with a three-page questionnaire. On the first page, they were asked to read a vignette that described a story with the key features noted in the earlier organized crime example. See Appendix A for the full text of the vignette.

On the second page, participants were asked to complete causal and counterfactual judgment tasks, the order of which was counterbalanced across participants. The causal question was phrased as follows: "Please list up to four factors that you regard as causes of Mr. Wallace's death." Four blank lines were printed below the question. After the causal listing task, participants were asked to "rate the importance of each factor you listed with regard to causality on a scale of 0–10 where 0 = *not at all causal* and 10 = *totally the cause*." Participants entered their ratings into boxes that were adjacent to each of the four lines. The counterfactual question was phrased as follows: "Please list up to four ways in which the story could be changed so that the outcome would have been different." Once again, participants were provided with four lines on which to list their responses. After the counterfactual listing task, participants were asked to "rate the importance of each of the changes you listed with regard to how likely that change would have been in altering the outcome on a scale of 0–10 where 0 = *not at all a good way to undo Mr. Wallace's premature death* and 10 = *absolutely the best way to undo Mr. Wallace's premature death*." Note that the "list-then-rate" approach for both causes and counterfactuals used in this experiment significantly improves on the common practice in similar studies (e.g., Mandel & Lehman, 1996; N'gbala & Branscombe, 1995; Wells & Gavanski, 1989) of having participants either list or rate counterfactuals and causes or of having participants list counterfactuals and rate causes.

On the final page, participants were asked to estimate four probabilities. Each question asked participants about the probability of the victim dying (a) "given that neither of the two specific events mentioned (*viz.*, Mr. Vincent adding poison to Mr. Wallace's drink and Mr. Bruce pushing Mr. Wallace's car into a ravine) had occurred," (b) "given that Mr. Vincent added poison to Mr. Wallace's drink but before Mr. Bruce pushed Mr. Wallace's car into a ravine," (c) "given that Mr. Bruce pushed Mr. Wallace's car into a ravine but ignoring that Mr. Vincent added poison to Mr. Wallace's drink," and (d) "finally, given that Mr. Vincent added poison to Mr. Wallace's drink and Mr. Bruce pushed Mr. Wallace's car into a ravine?" Participants responded by indicating a probability in percentages as an integer from 0 (*absolutely no chance of dying*) to 100 (*absolutely certain that he would die*).

Coding of causal and counterfactual listings. Three content categories were of a priori interest: statements focusing on CRIMELIFE, POISON, and CRASH. Two independent coders assigned participants' counterfactual and causal listings to these three categories or to a fourth "other" category. Each participant provided at least one counterfactual statement and at least

one causal statement that fell into the three content categories. The mean intercoder agreement across the three categories was 91%; discrepancies were resolved by the author.

Results

Causal and counterfactual judgments. In the present case, JDT predicts that CRIMELIFE will be the modal selection in participants' counterfactual listings and CRASH will be the modal selection in participants' causal listings. In contrast, CSA predicts that the modal selection in the causal listings will match the modal selection in the counterfactual listings. Moreover, these accounts make the same predictions about mean importance ratings. That is, JDT predicts a significant interaction between judgment type and target, such that participants will give the most weight to CRIMELIFE in their counterfactual ratings and the most weight to CRASH in their causal ratings. By contrast, CSA predicts that the target rated as most important in causal judgment will correspond with the target rated as most important in counterfactual judgment. I computed importance ratings by dummy coding participants' causal and counterfactual listings as 0 if the target was absent and as 1 if the target was present. The codes were then weighted by their corresponding importance ratings.³ Table 1 shows the percentage of participants listing each of the three target factors and the mean importance ratings for each of these factors as a function of judgment type. As can be seen in Table 1, the listings and ratings follow the same pattern. Therefore, I report inferential statistics only for the rating data, noting that the results for the listing data reveal the same pairwise effects.

To examine the predicted interaction effect, I conducted a 2 (judgment order) \times 2 (judgment type) \times 3 (target) mixed analysis of variance (ANOVA) on the importance ratings. The analysis revealed a significant main effect of judgment type, $F(1, 151) = 50.53$, $MSE = 10.00$, $p < .001$, partial $\eta^2 = .251$, and of target, $F(2, 302) = 27.24$, $MSE = 14.51$, $p = .001$, partial $\eta^2 = .153$. However, as JDT predicts, these effects were qualified by a significant Judgment Type \times Target interaction effect, $F(2, 302) = 136.60$, $MSE = 10.21$, $p < .001$, partial $\eta^2 = .475$.⁴ A series of planned comparisons further revealed that, in line with JDT, CRIMELIFE was rated as a more important counterfactual target than was either POISON, $t(151) = 4.41$, $p < .001$, Cohen's $d = 0.71$, or CRASH, $t(151) = 4.14$, $p < .001$, Cohen's $d = 0.67$. Conversely, CRASH was rated as a more important causal target than was POISON, $t(151) = 8.43$, $p < .001$, Cohen's $d = 1.37$, which, in turn, was rated as more important than CRIMELIFE, $t(151) = 9.53$, $p < .001$,

Cohen's $d = 1.55$. These results support JDT but are inconsistent with CSA.

CSA also predicts that the relation between counterfactual and causal judgments should be stronger when the counterfactual task precedes the causal task (Roese & Olson, 1997). To examine this "facilitation hypothesis," I calculated the mean within-target Pearson correlation as a function of judgment order. The mean correlation was .13 ($df = 151$, $p = .11$) when the causal task came first, and it was .18 ($df = 151$, $p = .03$) when the counterfactual task came first. Contrary to CSA, the difference between these correlations is not significant, $z = 0.44$, $p = .33$.

Probability judgment. JDT predicts a significant main effect of target on participants' probability judgment due to a sharp rise in the probability of the victim's death after the first sufficient condition for the victim's death occurs. As predicted, the main effect of target was significant, $F(3, 456) = 157.52$, $MSE = 1,216.33$, $p < .001$, partial $\eta^2 = .509$. Moreover, as Table 2 shows, the effect was due primarily to the steep rise from $P(\text{DEATH}|\text{CRIMELIFE})$ to $P(\text{DEATH}|\text{CRIMELIFE} \wedge \text{POISON})$, whereas the increase from $P(\text{DEATH}|\text{CRIMELIFE} \wedge \text{POISON})$ to $P(\text{DEATH}|\text{CRIMELIFE} \wedge \text{POISON} \wedge \text{CRASH})$ was relatively small. In fact, the change in outcome probability (ΔP) following POISON was significantly greater than the change following CRASH, paired $t(152) = 15.24$, SED (standard error of the difference) = 3.70, $p < .001$, Cohen's $d = 2.47$. On the basis of an assessment of these probability estimates, SPA predicts that, on average, participants would attach greater importance to POISON than to CRASH in their causal ratings. As noted in the previous subsection, however, the opposite result was observed. Finally, although mean causal importance ratings and mean probability change values did not show the same rank ordering, it is possible that there were significant positive correlations between these two measures for each target. However, these correlations were not reliable: $r(151) = -.14$, $p = .08$, for POISON, and $r(151) = -.04$, $p = .62$, for CRASH.

Discussion

Experiment 1 revealed a three-way dissociation among causal, counterfactual, and covariational judgments that confirmed the key predictions of JDT. First, counterfactual judgments focused primarily on the single event (CRIMELIFE) that would have been sufficient to prevent both the actual outcome (death by CRASH) and a similar, highly probable, yet counterfactual outcome (death by POISON). Second, covariational judgments focused on the first sufficient condition that yielded the largest increase in outcome probability. Third, participants attached the greatest causal importance to the sufficient condition (CRASH) that played a direct role in generating the actual outcome even though another sufficient condition (POISON) had already been intentionally enacted. The first two results support the substitution principle and the third result supports the actuality principle. Moreover, the pattern of findings

Table 1
Percentage of Participants Who Listed the Target and Mean Importance Ratings as a Function of Judgment Type (Experiment 1)

Target	Judgment type			
	Counterfactual		Cause	
	%	<i>M</i>	%	<i>M</i>
CRIMELIFE	75	5.37	31	2.17
POISON	53	3.27	81	5.77
CRASH	48	3.36	100	8.55

³ In cases where a target was listed more than once, multiple ratings for that target were averaged.

⁴ The main effect of target also was qualified by an interaction with judgment order, $F(2, 302) = 7.47$, $MSE = 10.21$, $p = .001$, partial $\eta^2 = .047$. Participants tended to attach greater importance to CRASH and less importance to CRIMELIFE and POISON when the causal questions were asked before rather than after the counterfactual questions.

Table 2
Mean Estimated Probability of Victim's Death in Percentages and Probability Change (ΔP) Values as a Function of Target (Experiment 1)

Target	Experiment 1		Replication	
	<i>P</i>	ΔP	<i>P</i>	ΔP
CRIMELIFE	19.05		26.29	
POISON	85.74	66.68	87.83	61.54
CRASH negating POISON ^a	85.78		83.73	
CRASH	95.98	10.25	98.22	10.39

^a This estimate is not required to compute the predictions of Spellman's probability-updating account but is useful in showing that CRASH was judged to be as effective as POISON in increasing the probability of DEATH ($t < 1$).

in Experiment 1 is inconsistent with the recency hypothesis (Miller & Gunasegaram, 1990), the primacy-constraint hypothesis (Wells et al., 1987), and the judgment-order effect predicted by CSA (Roese & Olson, 1997), all of which predict that the pattern of findings for causal judgment would match that for counterfactual judgment. The findings did not support SPA, either, which predicted that the first sufficient condition (POISON) would receive the most weight in causal judgment.

Replication Experiment

In Experiment 1, probability judgments were measured using an integer scale, whereas the causal judgments were based on the frequency of thought listings weighted by their corresponding importance ratings. To examine whether this difference in response format accounted for the disconfirmed predictions of SPA in Experiment 1, I conducted a replication experiment ($N = 40$ undergraduate volunteers) in which the causal importance of POISON and the causal importance of CRASH were rated on a Likert scale ranging from 0 (*not at all*) to 8 (*totally*). The probability questions were identical to those in Experiment 1. The results clearly replicated those of Experiment 1: First, CRASH ($M = 6.05$) was rated as a more important cause than was POISON ($M = 3.23$), paired $t(40) = 4.61$, $SED = 0.61$, $p < .001$, Cohen's $d = 1.46$. Second, the mean change in subjective probability of DEATH after POISON was significantly greater than the mean change after CRASH, paired $t(40) = 8.09$, $SED = 6.33$, $p < .001$, Cohen's $d = 2.56$ (see Table 2). Third, the within-target correlations between causal ratings and probability change scores were not reliable: $r(38) = .09$, $p = .58$, for POISON, and $r(38) = .03$, $p = .85$, for CRASH. The findings of the replication experiment lend additional support to JDT and do not support SPA.

Experiment 2

The results of Experiment 1 clearly demonstrate that causal, counterfactual, and covariational reasoning can diverge in focus in cases where the antecedents that best maximize the main goal of each type of reasoning also differ. Nevertheless, support for JDT would be strengthened considerably if it was shown that a consistent pattern of findings could be obtained using a different case that also differed in terms of various features (e.g., no intentional

wrongdoing among the actors involved, longer time duration, more turning points in the story). This was the primary objective of Experiment 2. The new vignette described a case with the following key features: A man (John) faced a decision of whether to change jobs, and he eventually opted for the change. Consequently, over most of the next 30-year period, he was exposed to asbestos at work. This caused him to develop lung cancer, from which he was sure to die in the coming months. On the day of John's death, he had a routine visit to the hospital. The nurse on call accidentally administered the wrong medication, to which John happened to be allergic. John died moments later from a heart attack and shock induced by the drug.

Three predictions were made. First, participants would be more likely to focus on John's decision to change jobs in their counterfactual judgments than in their causal judgments because by undoing that decision, John's inevitable death by lung cancer and his actual death due to medical error would have been avoided. Second, the largest increase in the rated subjective probability of John's death would occur after participants updated that probability given the exposure to asbestos, which represents the earliest sufficient condition for John's death in this case. Third, participants would be more likely to focus on the nurse's misadministration of the medication in their causal judgments than in their counterfactual judgments because this factor played a direct role in how John's actual death was brought about. The first two predictions follow from the substitution principle; the last follows from the actuality principle. An additional objective of Experiment 2 was to test more directly than in past research (e.g., Mandel & Lehman, 1998, Experiment 1) the idea that people think about causes primarily in terms of sufficiency rather than necessity. Participants were asked to choose whether the expression "X causes Y" means "when X happens, Y also will happen" (i.e., X is sufficient for Y) or "when X doesn't happen, Y also won't happen" (i.e., X is necessary for Y). If CSA is accurate, then a significantly greater proportion of participants should choose the necessity definition. In contrast, if causation is understood primarily in terms of sufficiency, then the opposite tendency should be observed.

Method

Participants. Fifty-eight undergraduate students participated in this experiment for optional extra credit points in an introductory psychology course.

Materials and procedure. Participants were presented with a four-page questionnaire. On the first page, participants read that "the aim of this study is to examine how people think about different types of events. Please begin by carefully reading the following story about 'John's Misfortunes.'" The vignette that followed is included with this article as Appendix B.

On the second page, participants were asked to complete the causal and counterfactual judgment tasks, the order of which was counterbalanced across participants. The causal question was phrased as follows: "Given all that you know about the events just described, who or what caused John's premature death? Please list up to four factors that you regard as the cause or causes of John's premature death." Four blank lines were printed below the question. After the causal listing task, participants were asked to "rate the importance of each factor you listed with regard to causality on a scale of 0–10 where 0 = *not at all causal* and 10 = *totally the cause*." The counterfactual question was phrased as follows: "Given all that you know about the events just described, please list up to four ways the story could be changed so that John's premature death would have been avoided." Once again, participants were provided with four lines on which to list their

responses. After the counterfactual listing task, participants were asked to “rate the importance of each of the changes you listed with regard to how likely that change would have been in altering the outcome on a scale of 0–10 where 0 = *not at all a good way to undo John’s premature death* and 10 = *absolutely the best way to undo John’s premature death*.”

On the third page, participants were asked to estimate five probabilities. Each question asked “What do you think John’s chances were of dying prematurely,” conditional on the following significant points in the story: (a) “at the point at which John was first hired at the manufacturing company at the age of 17?” (b) “at the point at which John accepted the offer from the plastics division?” (c) “at the point at which John was first moved to the new location with asbestos?” (d) “at the point at which John had spent 30 years in the building with asbestos?” and (e) “at the point at which John was administered the wrong medication by the nurse?” As in Experiment 1, participants provided their responses by indicating a probability in percentages as an integer from 0 (*absolutely no chance of dying*) to 100 (*absolutely certain that he would die*).

Finally, on the last page, participants were asked two questions that examined their conceptions of causation. The first question was “Which of the following statements best describes what *you* mean by the phrase ‘X causes Y’?” The second question was “Which of the following best describes what *most people* mean by the phrase ‘X causes Y’?” The response options for each question were (a) “when X happens, Y also will happen” and (b) “when X doesn’t happen, Y also won’t happen.” As noted above, these options correspond to sufficient cause and necessary cause interpretations, respectively.

Coding of causal and counterfactual listings. Seven content categories were of a priori interest: statements focusing on (a) the availability of a job at the manufacturing firm when John was 17 years old (JOB), (b) John’s decision to change jobs (DECISION), (c) the plastics division’s relocation (MOVE), (d) the long-term exposure to asbestos and/or the resulting cancer (CANCER), (e) the nurse’s carelessness and/or misadministration of medication (DRUG), (f) the allergic reaction and/or resulting heart attack and shock (REACTION), and (g) the failed emergency response (RESPONSE). The first five categories correspond to the five probability questions. The latter two were added to examine whether participants have a preference for focusing on proximal (DRUG) or immediate (REACTION and RESPONSE) causes. Two independent coders assigned participants’ counterfactual and causal listings to these seven categories. Each participant provided at least one counterfactual statement and at least one causal statement that fell into the seven content categories. The mean intercoder agreement across the seven categories was 92% (93% agreement for the causal listings and 90% for the counterfactual listings). Discrepancies were resolved by the author. As in Experiment 1, I dummy coded participants’ listings as 0 if the target was absent and as 1 if the target was present. These codes were then weighted by their corresponding importance ratings, thus providing a possible range from 0 to 10.

Results

Causal and counterfactual judgments. Table 3 shows the percentage of participants who listed each of the seven targets and the corresponding mean importance ratings. Given the similar pattern of results for the listings and ratings, I present inferential analyses for ratings only. As predicted by JDT, the interaction of target and judgment type was significant, $F(6, 330) = 18.00$, $MSE = 6.14$, $p < .001$, partial $\eta^2 = .247$.⁵ To examine the nature of this interaction, I conducted a series of planned comparisons, which revealed strong support for JDT: First, DECISION was rated as more important in counterfactual judgments than in causal judgments, $t(57) = 6.87$, $SED = 0.59$, $p < .001$, Cohen’s $d = 1.82$. Second, DRUG was rated as more important in causal judgments than in counterfactual judgments, $t(57) = 3.54$, $SED = 0.57$, $p < .001$,

Table 3
Percentage of Participants Who Listed the Target and Mean Importance Ratings as a Function of Judgment Type (Experiment 2)

Target	Judgment type			
	Counterfactual		Cause	
	%	<i>M</i>	%	<i>M</i>
JOB	7	0.60	0	0.00
DECISION	79	5.66	38	1.59
MOVE	9	0.78	7	0.38
CANCER	52	3.97	66	5.03
DRUG	71	5.36	95	7.38
REACTION	9	0.28	19	1.53
RESPONSE	3	0.23	2	0.09

Cohen’s $d = 0.94$. Third, DRUG was rated as a more important cause than CANCER, $t(57) = 3.24$, $SED = 0.73$, $p < .002$, Cohen’s $d = 0.86$, which in turn was rated as more important than DECISION, $t(57) = 5.49$, $SED = 0.63$, $p < .001$, Cohen’s $d = 1.45$. Finally, although DECISION was the most important counterfactual target and was rated as significantly more important than CANCER, $t(57) = 2.41$, $SED = 0.70$, $p < .02$, Cohen’s $d = 0.64$, it did not significantly differ from DRUG, $t < 1$. However, the latter result is neither particularly surprising nor troubling for JDT given the highly abnormal and apparently controllable nature of the nurse’s error. As previous research has shown, counterfactuals tend to focus on antecedents that are abnormal (Kahneman & Tversky, 1982a) and controllable (Giroto, Legrenzi, & Rizzo, 1991; Mandel & Lehman, 1996; McCloy & Byrne, 2000). Finally, the mean correlation between causal and counterfactual importance ratings calculated across all targets except those with few or no responses (viz., JOB and REACTION) was .23 ($p = .08$) when the counterfactual task came first and .13 ($p = .33$) when the causal task came first. Consistent with the results of Experiment 1 and contrary to CSA, the difference between these mean correlations was not reliable, $z = 0.37$, $p = .36$.

Probability judgments. The substitution principle predicts that the largest change in outcome probability would occur after updating on MOVE and/or CANCER. Table 4 shows the mean probability judgments and the corresponding ΔP values. Confirming JDT’s prediction, the largest change in subjective probability was associated with the exposure to asbestos. Approximately half of that change in probability was associated with MOVE ($\Delta P = 23.28$) and half with CANCER ($\Delta P = 22.17$). In contrast, updating on DRUG led to only a 3% change in outcome probability. Given these results, SPA predicts that most participants would select either MOVE or CANCER as the cause of death. However, as reported above—and in support of the actuality principle—DRUG was rated a more important causal factor. Finally, consistent with Experiment 1 and contrary to SPA, the correlations between causal ratings and probability change values for DECISION, MOVE, CANCER, and DRUG all were

⁵ As in Experiment 1, the main effect of target was significant, $F(6, 330) = 67.65$, $MSE = 10.30$, $p < .001$, partial $\eta^2 = .552$, as was the Target \times Judgment Order interaction effect, $F(6, 330) = 5.40$, $MSE = 10.30$, $p < .001$, partial $\eta^2 = .089$.

Table 4
Mean Estimated Probability of Death in Percentages and Probability Change (ΔP) Values as a Function of Target (Experiment 2)

Target	<i>P</i>	ΔP
JOB	29.32	
DECISION	40.39	11.07
MOVE	63.67	23.28
CANCER	85.84	22.17
DRUG	88.96	3.12

unreliable: $r_s(56) = -.10, -.12, -.04,$ and $.06$ (smallest $p = .37$), respectively.

Concepts of causation. In line with earlier sufficiency-focused accounts of causal understanding (Mandel & Lehman, 1996, 1998), JDT predicts participants' modal definition of causation would focus on the sufficiency interpretation provided as one of the two response options. Confirming this prediction, 81% of participants indicated that they defined causation in terms of sufficiency ("if X happens, then Y also will happen") and 84.5% of participants indicated that most people would define causation in this manner (binomial $ps < .001$).

Discussion

Experiment 2 replicated the pattern of results obtained in Experiment 1 while using a different, more complex case. This yielded the three-way dissociation among causal, counterfactual, and covariational judgments predicted by JDT. First, counterfactual judgments focused primarily on the controllable event (DECISION) that would have been sufficient to prevent both the actual outcome (death due to DRUG) and a similar, highly probable, yet counterfactual outcome (death due to MOVE and CANCER). Second, covariational judgments focused on the first sufficient condition that yielded the largest increase in outcome probability (MOVE and CANCER). Third, participants attached the greatest causal importance to the sufficient condition (DRUG) that played a direct role in generating the actual outcome even though another process sufficient to yield premature death (CANCER) had already been set in motion. This dissociation in judgment focus is inconsistent with both CSA and SPA, but it supports the key predictions of JDT, namely, that counterfactual and covariational reasoning conform to the substitution principle and that causal reasoning conforms to the actuality principle. Note that the findings also supported a modified version of the primacy-constraint hypothesis (Vinokur & Azjen, 1982; Wells et al., 1987). Although participants showed a preference for listing the later sufficient cause (DRUG) that directly contributed to the actual cause of death, few participants focused on the immediate causes of death (*viz.*, REACTION and RESPONSE). A modified version of the primacy-constraint hypothesis that does not assume counterfactual reasoning has a key role (*cf.* Wells et al., 1987) and that conforms to the actuality principle lends further specification to JDT.

Finally, the present findings directly supported the idea that people understand causation in terms of sufficient rather than necessary causes. These findings are consistent with other recent findings that suggest that people think about causes in terms of

sufficiency (Mandel & Lehman, 1996, 1998; see also Goldvarg & Johnson-Laird, 2001), and they provide the most direct evidence to date for this proposal.

Experiment 3

In Experiment 3, I revisit the coin-toss problem with the goal of pitting SPA against JDT using a methodology that differs from that used in Experiments 1 and 2. A basic finding from research using this problem (Byrne et al., 2000; Miller & Gunasegaram, 1990; Spellman, 1997) is that participants tend to predict that Player 1 will ascribe more blame to Player 2 than vice versa. As noted above, SPA posits that this finding is due to the fact that the probability of the outcome (win \$0) does not change after Player 1's coin toss but increases from .50 to certainty after Player 2's coin toss. In each of the studies just cited, however, participants were asked to judge which player would blame the other player more. This represents a prediction task in which the participant is asked to imagine the attributional perspective of each player with regard to the other and then to make a comparative judgment based on the dual forecasts. In contrast, SPA is intended to describe how people attribute causation (or blame) to actors or events in a case (on the relation between attributions of blame and causation, see Shaver, 1985). Therefore, it would be more relevant to ask participants how they personally would ascribe blame to the two players. If SPA is accurate, participants should ascribe more blame to Player 2 than to Player 1 in spite of the fact that from a normative perspective, given the independence of the each player's result, both players contributed equally to the outcome.

Alternatively, if, as JDT posits, participants are sensitive to the process leading to the outcome—namely, in the present case, that the rule of the game demanded a particular type of joint outcome to which both players contributed equally via a chance process—then the modal response should be to ascribe blame equally to the two players, in line with normative expectation. Two of the studies in which researchers examined this problem (Byrne et al., 2000; Miller & Gunasegaram, 1990) provided participants with tasks that biased their responses in favor of choosing one player or the other. Miller and Gunasegaram (1990) gave participants a forced choice between Players 1 and 2. Byrne et al. (2000) used a free-response format but asked the question "Who will blame the other more—Jones or Brady?" (p. 266). Clearly, this phrasing suggests that participants should choose one or the other player to be identified as more blameworthy. Spellman (1997) provided participants with three response options (Player 1, Player 2, or the same) but then summarized the findings in a manner that likely underestimated the frequency of "same" responses. Namely, if at least one response to the multiple questions posed (*i.e.*, about blame, counterfactuals, and guilt) focused on one player—say Player 2—and none of the remaining responses focused on the other player, then the participant's overall response was coded as "Player 2." In contrast, participants coded as responding "the same" had to respond this way to all three questions. Therefore, none of the extant studies has tested definitively whether participants attribute more blame to Player 2 than to Player 1. Experiment 3 tested JDT's prediction that most participants, in fact, would ascribe blame equally to the two players when the preceding methodological limitations were eliminated.

Experiment 3 also introduced a new problem in which the rule for winning is that both coins must land on tails. However, both players' coins land on heads, so they win nothing. In this case, the probability of the outcome (win \$0) is .75 at the start of the game, and it rises to certainty after Player 1's coin toss. Player 2 does not change the probability. Thus, SPA predicts that the modal response to the blame question will be Player 1. The "both-tails" problem permits the prediction of SPA to diverge from Miller and Gunasegaram's (1990) recency hypothesis, which predicts that the modal response will be Player 2. In contrast to both accounts, JDT predicts that the modal response will be that the amount of blame is the same, because the rule requires a specific outcome of each player's coin toss that, in fact, neither player produced. It is also of interest to compare the percentage of participants providing a response of "the same" across the two problems. In the both-tails problem, each player yields an outcome that is sufficient to ensure that they do not win. In contrast, in the same-outcome problem, neither player yields an outcome that on its own is sufficient to ensure that the players do not win. Rather, it is the combination of their outcomes that is sufficient. Given this difference in process, only Player 2 could know what outcome to aim for in the same-outcome case, whereas both players know exactly what to aim for in the both-tails case. Accordingly, JDT predicts that the percentage of "the same" responses should be even greater in the both-tails case than the same-outcome case to the extent that participants believe the players would have some control over the outcomes of their coin tosses.

Method

Participants and design. Fifty-six undergraduate students participated on a voluntary basis. Using a one-way between-subjects design, I randomly assigned participants in equal numbers to one of two problem conditions (same outcome vs. both tails).

Materials and procedure. The same-outcome problem was described as follows:

Imagine two individuals (Jones and Cooper) who are offered the following very attractive proposition. Each individual is asked to toss a coin. If the two coins come up the same (both heads or both tails), each individual wins \$1,000. However, if the two coins do not come up the same, neither individual wins anything. Jones goes first and tosses a head. Cooper goes next and tosses a tail. Thus, the outcome is that neither individual wins anything.

The both-tails problem was similar except that the rule for winning was "If the two coins both come up tails, each individual wins \$1,000. However, if either or both coins land on heads, neither individual wins anything." And each player's coin landed on heads.

After reading the problem, participants were asked two blame questions. The first involved forecasting which player would experience more self-blame ("Who do you predict will experience more self-blame?"). This question, although focusing on self- rather than other-blame, was akin to the question posed in previous studies in the sense that it required participants, as observers, to make attributional forecasts from the actors' perspectives and then compare them to decide which is more extreme. The second question involved a direct attributional judgment: "Who do you think is more to blame for the failure to win the prize?" For both questions, participants were presented with a choice between three response options: (a) "Jones," (b) "Cooper," or (c) "the same."

Results

Table 5 presents the percentage of participants who provided each type of response as a function of problem (same outcome vs. both tails) and perspective (player vs. participant).

Same-outcome problem. Consistent with past research (Byrne et al., 2000; Miller & Gunasegaram, 1990; Spellman, 1997), a majority (68%) of participants forecasted that Player 2 (Cooper) would experience more self-blame than Player 1 (Jones) in the same-outcome condition. This percentage was marginally greater than the percentage of participants (32%) who forecasted that self-blame would be the same for the two players, binomial $p = .09$. In contrast, when blame was attributed to the two players from the participants' perspective, a very different finding was obtained. Supporting the prediction of JDT and disconfirming the prediction of SPA, the vast majority (86%) of participants indicated that the two players were equally blameworthy. This percentage was reliably greater than the percentage of participants who found one player to be more blameworthy than the other (14%), binomial $p < .001$. Moreover, it also was reliably greater than the percentage of participants who forecasted equal levels of self-blame, McNemar test $p < .001$, partial $\eta^2 = .536$. The last result is important because both JDT and SPA are primarily designed to account for first-order judgments, not higher order forecasts of actors' judgments.

Both-tails problem. In the both-tails condition, JDT once again predicts that participants will attribute blame equally to the two players, especially when the blame attributions are made from the participant's perspective rather than the player's perspective. In contrast, SPA predicts that Player 1 should be judged as most blameworthy. The results provided strong support for JDT and were inconsistent with the predictions of SPA. First, a significantly greater percentage of participants indicated that self-blame would be attributed equally (75%) rather than unequally (25%) to both players, binomial $p < .02$. Second, all but one participant (96%) attributed blame equally to the two players, binomial $p < .001$. Third, the percentage of participants responding "the same" was reliably greater when answered from the participant's perspective (96%) than from the player's perspective (75%), McNemar test $p < .04$, partial $\eta^2 = .214$.

Comparison of responses across problems. Finally, to test JDT's prediction that the percentage of participants who responded "the same" would be significantly greater in the both-tails condition than in the same-outcome condition, I computed a variable in which participants who did not respond "the same" on either blame question were given a score of 1, participants who responded "the

Table 5
Percentage of Participants Indicating a Given Response as a Function of Problem and Perspective (Experiment 3)

Response	Problem			
	Same outcome		Both tails	
	Player	Participant	Player	Participant
Player 1	0.0	3.6	17.9	3.6
Player 2	67.9	10.7	7.1	0.0
Same	32.1	85.7	75.0	96.4

same” on only one question were given a score of 2, and participants who responded “the same” on both questions were given a score of 3. An analysis of the rank order of scores as a function of problem confirmed JDT’s prediction that participants were more likely to provide a response of “the same” in the both-tails condition (mean rank = 34.66) than in the same-outcome condition (mean rank = 22.34), Mann-Whitney $U = 219.50$, $p < .003$, partial $\eta^2 = .169$.

Discussion

The findings of Experiment 3 strongly supported JDT but did not support SPA. Across all responses generated in Experiment 3, 72% matched the response predicted by JDT, 25% matched the response predicted by SPA, and the remaining 3% did not match the predictions of either account. The findings also challenge the claim by Miller and Gunasegaram (1990) that participants are more likely to ascribe blame to actors whose behaviors come later in a temporal chain of events than to actors whose behaviors come earlier because the actions of the former are more mutable than those of the latter. If that were so, one would expect most participants to have attributed blame to Player 2 in both problems. In fact, however, very few participants attributed blame to Player 2 in the newly constructed problem despite the fact that it was a noncausal chain (i.e., Player 2’s outcome was independent of Player 1’s outcome). Finally, Experiment 3 highlights the importance of providing participants with the full set of response options and unbiased questions. Past research using this problem has not recorded participants’ true modal response because either (a) the option of attributing blame equally to the two players was not provided, (b) participants were presented with leading questions that conveyed the idea that blame should be ascribed to only one player, or (c) the data were summarized in a biased manner.

Experiment 4

Experiment 4 had two objectives. First, I examined whether participants’ tendency to respond “the same” in the standard (same-outcome) version of the coin-toss problem would generalize from blame judgments to causal judgments. In Experiment 3, participants were asked to provide blame judgments to increase the comparability of the findings to past research that has used this problem. However, given that the focus of this article is on causal selection, I tested whether the findings were replicable using a causal selection task. Second, Experiment 4 directly tested whether the substitution principle characterizes participants’ covariational judgments. According to the substitution principle, participants will interpret the outcome of the game in terms of the ad hoc category *not winning the prize* rather than the actual outcome of Player 1 tossing heads and Player 2 tossing tails. Participants asked to judge the post hoc probability of the outcome from the start of the game should therefore provide estimates close to .50 (i.e., two out of the four possibilities leading to no prize—viz., tails–heads and heads–tails). In contrast, if participants do not adopt this category-based focus, they would be expected to provide estimates closer to .25 because there is a 1-in-4 chance of obtaining the heads–tails outcome from the outset.

Method

Participants and design. Forty-three undergraduate students participated and received optional extra credit points in an introductory psychology course. Participants were randomly assigned to one of two conditions that differed only in terms of whether the causal task preceded the probability tasks ($n = 21$) or whether the probability tasks came first ($n = 22$).

Materials and procedure. Participants were given a questionnaire that first described the same-outcome version of the coin-toss problem described in Experiment 3. Next, participants were presented with one causal judgment task and three probability estimation tasks. The probability estimation tasks were presented in the fixed order described below. Each question began with the phrase “Right from the start of the game, what was the probability. . .” The three questions ended as follows: (1) “of the game’s outcome?” (2) “that Jones and Cooper would not win prizes?” and (3) “that the coin would land on heads when Jones tossed it and that it would land on tails when Cooper tossed it?” Thus, the first question left the outcome ambiguously defined, whereas the second question adopted an explicit category-based focus and the third question focused on the actual outcome. Participants responded by indicating a probability in percentages as an integer from 0 (*absolutely no chance at all*) to 100 (*absolutely certain*). For the causal judgment task, participants were asked,

Which of the following statements best represents your belief about the cause of the game’s outcome?

- (a) Jones is more the cause of not winning the prize than is Cooper.
- (b) Jones and Cooper are about equally the cause of not winning the prize.
- (c) Cooper is more the cause of not winning the prize than is Jones.

The order of the causal and probability questions was counterbalanced across participants.

Results and Discussion

Causal judgment. As noted above, JDT predicts that participants will attribute causation equally to Players 1 (Jones) and 2 (Cooper), whereas SPA predicts that most participants will select Player 2 as the cause. Consistent with the results of Experiment 3, 91% of participants attributed causation to the two players equally, with the remaining 9% attributing causation to Player 2, binomial $p < .001$. Therefore, Experiment 4 provided additional support for JDT by showing that the findings obtained in Experiment 3 for the same-outcome version of the coin-toss problem generalize from blame judgments to causal judgments.

Probability judgment. The substitution principle predicts that the modal response to the ambiguous-outcome question would match the modal response to the explicit category-based-outcome question and that the estimate provided should be close to 50 (i.e., $P = .5$). In contrast, the modal response to the actual-outcome question was predicted to be closer to 25 (i.e., $P = .25$). Matching this prediction perfectly, the modal responses to the ambiguous-outcome and category-based-outcome questions were 50 in each case, whereas the modal response to the actual-outcome question was 25. The median probability estimates were identical to the modal estimates. Therefore, Experiment 4 provides direct evidence in support of the substitution principle. Participants spontaneously interpreted the outcome of the game to mean “not winning the prize” rather than the actual heads–tails outcome.

General Discussion

The results of Experiments 1–4 all provide support for JDT. In Experiments 1 and 2, the predicted three-way dissociations among

causal, counterfactual, and covariational judgments were observed. In support of the substitution principle, participants placed the greatest emphasis in counterfactual judgment on antecedents whose negation was sufficient to prevent not only the actual outcome of the case but also another counterfactual outcome that was functionally similar to the actual outcome and that was almost certain to occur in the future. Moreover, participants judged that the largest increase in outcome probability was conditional on antecedents that virtually guaranteed an outcome that was functionally similar to but played no direct role in generating the actual outcome. Experiment 4 provided direct evidence in support of the substitution principle by showing that participants spontaneously define the outcome of a case in category-based terms when making post hoc assessments of its probability. In support of the actuality principle, participants in Experiments 1 and 2 placed the greatest emphasis in causal judgment on antecedents that played a direct role in how the actual outcome was generated. Moreover, the related JDT proposal that causal reasoners are sensitive to the processes by which a given outcome is generated was supported by the findings of Experiment 3, which showed that the vast majority of participants attributed blame to hypothetical actors in accordance with the rules of the game that they were asked to consider. Finally, building on past research (Goldvarg & Johnson-Laird, 2001; Mandel & Lehman, 1998), Experiment 2 provided direct support for the idea that people understand the meaning of causation primarily in terms of sufficiency (i.e., X is sufficient to generate Y) rather than necessity (i.e., X is necessary to generate Y).

Experiments 1–4 each provided a strong inferential test (Platt, 1964) of at least two competing theories. In Experiments 1 and 2, the predictions of JDT were pitted against those of CSA and SPA. Contrary to the predictions of CSA, the ranked mean importance of antecedents in causal and counterfactual judgment did not coincide, and the effect of judgment order on the correlations between causal and counterfactual judgments was negligible. The findings of these experiments are also inconsistent with the primacy-constraint (Wells et al., 1987) and the recency (Miller & Gunasegaram, 1990) hypotheses, which predict that causal judgment will focus on the most mutable antecedent. Contrary to the predictions of SPA, the antecedents that were judged by participants to have led to the largest increase in outcome probability were not the ones that participants mentioned most frequently or gave the most weight to in causal judgment. Further, causal importance ratings and probability change values were not positively correlated in either experiment. Experiments 3 and 4, in which I utilized a key problem that Spellman (1997) used to demonstrate support for SPA, further revealed that when past methodological limitations were corrected, the findings supported JDT but not SPA. Experiment 3 also directly disconfirmed the recency hypothesis.

The Functionalist Stance of JDT

As noted above, JDT proposes that causal, counterfactual, and covariational reasoning each serve an adaptive learning function. Although the functional objectives of explanation, prediction, and control generally apply to each of these three types of reasoning, the focus of what is to be learned differs somewhat for each type. By diversifying the foci of causal, counterfactual, and covariational reasoning, reasoners are likely to arrive at a much fuller description of the critical and decisive events in a relevant case. In

many cases, causal, counterfactual, and covariational selections will focus on the same events because those events will have implications for multiple functional objectives. For example, if an arsonist sets a house ablaze and there was little anyone else could do to stop him, then all three types of judgment may focus on the arsonist's actions and motives. In contrast, the cases examined in Experiments 1 and 2 were designed to tease apart the different foci of judgment by incorporating a series of antecedents that would, according to the predictions of JDT, be weighted differently by participants using the three corresponding types of reasoning. In the following subsections, I expand on JDT's functionalist stance.

Causal reasoning. JDT posits that causal reasoning focuses on understanding or explaining the processes that produced a particular outcome. The actuality principle is based on this idea. Accordingly, JDT shares with other mechanism accounts the view that causal reasoning is focused on answering the question of how the outcome came to be. As Ahn and Kalish (2000) have noted, the emphasis on *how* differs from covariational accounts, which almost invariably question participants about who or what they select as the cause. For example, in Kelley's (1967) ANOVA model, causal reasoners use covariational information about consensus, stimulus distinctiveness, and situational consistency to attribute causation to either the person, the stimulus, or the occasion. But to say, for example, that X happened because of something about the actor begs the question: What was it about the selected cause that led to X? In fact, Ahn et al. (1995) demonstrated that people are more likely to seek information about causal mechanism rather than covariation. The present research builds on Ahn et al. (1995) by showing that causal selection focuses on antecedents that provided a good answer to the question of how an outcome came to be, even when the selected antecedents are less important than others from a probability-updating perspective. It is interesting to note that in neither Experiment 1 nor Experiment 2 was the task phrased as a *how* question, suggesting that participants spontaneously adopted a *how* focus. However, JDT assumes that the focus of a causal question may shift to accommodate the goals of reasoning. For example, recent research (McClure, Hilton, Cowan, Ishida, & Wilson, 2001) demonstrated that if the means to an end are known or believed to be easily obtainable, then it is more informative to pose a *why* question, prompting an inquiry of the focal actor's goals.

Counterfactual reasoning. JDT posits that counterfactual reasoning serves a somewhat different adaptive learning function. Namely, the theory posits that counterfactual reasoning focuses on identifying ways that would have been sufficient under the circumstances to prevent the actual outcome or something like it from occurring (cf. Mandel & Lehman, 1996). Accordingly, counterfactual selections should satisfy the goal of identifying ways of undoing not only the actual outcome but also other outcomes that are similar in terms of their consequences, especially if at some earlier point in the sequence of events the relevant outcome had a high propensity to occur (Kahneman & Varey, 1990). This was the case in Experiments 1 and 2, in which participants tended to list and give the most weight to antecedents that would undo not only the actual outcome but also a similar outcome that was likely to occur in the future had the causal process culminating in the actual outcome not occurred. JDT can also account for recent findings of Spellman and Kincannon (2001). Participants read a vignette in which two assassins shot a victim simultaneously, and each shot

was sufficient to kill the victim. Although a majority of participants attributed causation to each assassin individually, participants were three times more likely to undo both assassins' actions (68%) rather than either one's individual actions (23%) in their counterfactual listings. If participants had focused on undoing only the actual outcome (death by two shots), then mentally negating either shot would have worked. However, if participants had focused on undoing a category of outcomes (e.g., that in which the protagonist is killed on that day), then both shots would have had to be negated. The possible world in which one shot was fired appears to be too similar to the actual world in terms of its outcome to constitute a compelling case of undoing.⁶ In summary, JDT proposes that ad hoc categories play a much more important role in counterfactual reasoning than past accounts have assumed.

Covariational reasoning. Although SPA proposes that covariational judgments play an important role in causal selection, it does not specify how those judgments are made. In contrast, JDT posits that such judgments will tend to conform to the substitution principle. As Experiments 1 and 2 demonstrated, the greatest increase in the subjective probability of the outcome followed events that in fact did not play a direct role in generating the actual outcomes. Rather, the antecedents that were seen to make the biggest difference would have led to functionally similar outcomes by completely different causal processes. Experiment 4 demonstrated even more directly that participants asked to estimate the probability of the outcome interpreted the outcome in terms of the ad hoc category *not winning the prize* rather than in terms of ways of getting the actual heads-tails outcome. As with counterfactual reasoning, there is a functional basis for this broader perspective in covariational assessments. By focusing on a category of functional significance, reasoners can assess the likelihood of a certain type of consequential outcome occurring independent of the specific way in which it occurred. Moreover, in cases involving multiple sufficient conditions for a type of consequential outcome, mental substitution allows reasoners to identify the earliest point in a chain of events whereby that type of outcome became either probable or necessary.

Implications for Prescriptive Theories of Causal Reasoning

In cases involving multiple sufficient causes, SPA predicts that reasoners will discount all but the earliest sufficient cause. It is clear that in some cases, such as the coin-toss problem examined in Experiments 3 and 4, this pattern of discounting is unwarranted and would appear to reflect an irrational aspect of human judgment (Miller & Gunasegaram, 1990). That is, each player has the same likelihood of producing one result or the other, the two results are independent, and each player's result has the same weight in determining the overall outcome. As these experiments showed, most participants attributed blame and causation equally to the two players. Therefore, most judgments were normative. Moreover, the findings of Experiments 1 and 2 demonstrated that participants' causal importance judgments did not reflect the type of asymmetric discounting predicted by SPA. According to JDT, the discounting of sufficient conditions that do not play a role in the process by which the actual outcome of a case was generated is prescriptively valid because the goal of causal reasoning is to learn about causal mechanisms, processes, or constraints that might be projectable to

future cases. Given that human rationality is bounded by cognitive constraints, the ability to apply such knowledge to future cases represents a highly efficient learning strategy. This is also an important reason why strictly bottom-up covariational accounts of causal reasoning lack psychological plausibility.

Directions for Future Research

Unlike previous accounts, JDT proposes that ad hoc categorization (Barsalou, 1983, 1991) plays an important role in counterfactual and covariational reasoning. Future research might profitably examine the factors that constrain outcome categorization in the processes of undoing and of postdiction. One intriguing question concerns the ways in which a simulated outcome must differ from the actual outcome for it to be viewed as a successful example of undoing. Similarly, in what ways must an outcome judged to be highly probable given a particular antecedent have occurred resemble the actual outcome of the case? For example, would an event that would "merely" have led to a severe injury also need to be undone in the process of successfully undoing a fatality generated by a later causal process? Such questions open up the possibility for studies linking research on reasoning, judgment, and categorization.

Another important objective of future research will be to examine how the predictions of JDT and other competing accounts fare as a function of changes in context. Mandel (2003a) has shown that the frequency and type of counterfactuals and attributions generated by participants recalling significant life events varied as a function of situational context. In the present research, the cases considered all involved negative outcomes. Perhaps the focus on prevention in counterfactual reasoning would shift toward promotion if the outcomes were positive (Roese, Hur, & Pennington, 1999). Future studies could examine how outcome valence and other case characteristics influence each of the three types of judgment processes examined in this article.

Finally, future research could examine whether forecasters and planners benefit from reasoning about a problem from different reasoning perspectives. For instance, how (and also how well) do planners weight and combine the knowledge that they obtain from these different reasoning perspectives? Building on earlier work (e.g., Roese, 1994), future research could compare the performance and calibration of confidence among individuals who are directed to think either causally, counterfactually, or covariationally with results from others who are directed to use multiple types of reasoning to arrive at a forecast, plan, or decision.

⁶ It is interesting to note that this finding also suggests that the goal of undoing the appropriate ad hoc category of outcomes overrides the *minimal rewrite rule* (Tetlock & Belkin, 1996), which states that counterfactuals intended to be plausible tend to include a minimal number of changes, and the changes that are made preserve a high degree of closeness between the actual and counterfactual cases.

References

- Ahn, W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31*, 82-123.
- Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal

- reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–225). Cambridge, MA: Cambridge University Press.
- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*, 299–352.
- Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition*, *23*, 510–524.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.
- Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 27, pp. 1–64). New York: Academic Press.
- 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.
- Brewer, M. B. (1977). An information-processing approach to attributions of responsibility. *Journal of Experimental Social Psychology*, *13*, 58–69.
- Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209–254). New York: Academic Press.
- Byrne, R. M. J., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000). The temporality effect in counterfactual thinking about what might have been. *Memory & Cognition*, *28*, 264–281.
- Cheng, P. W. (1993). Separating causal laws from causal facts: Pressing the limits of statistical relevance. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 215–264). New York: Academic Press.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.
- Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, *40*, 83–120.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago: Rand McNally.
- Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory & Cognition*, *23*, 646–658.
- Davis, C. G., Lehman, D. R., Wortman, C. B., Silver, R. C., & Thompson, S. C. (1995). The undoing of traumatic life events. *Personality and Social Psychology Bulletin*, *21*, 109–124.
- Fairley, N., Manktelow, K., & Over, D. (1999). Necessity, sufficiency, and perspective effects in causal conditional reasoning. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *52A*, 771–790.
- Giroto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, *78*, 111–133.
- Goldvarg, E., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565–610.
- Greene, E. J., & Darley, J. M. (1998). Effects of necessary, sufficient, and indirect causation on criminal liability. *Law and Human Behavior*, *22*, 429–451.
- Harré, R., & Madden, E. H. (1975). *Causal powers: A theory of natural necessity*. Totowa, NJ: Rowman & Littlefield.
- Hart, H. L. A., & Honoré, A. M. (1985). *Causation in the law* (2nd ed.). Oxford, England: Clarendon Press.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Hilton, D. J. (1988). Logic and causal attribution. In D. J. Hilton (Ed.), *Contemporary science and natural explanation: Commonsense conceptions of causality* (pp. 33–65). New York: New York University Press.
- Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, *107*, 65–81.
- Hilton, D. J., & Erb, H.-P. (1996). Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking & Reasoning*, *2*, 273–308.
- Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. New York: Basic Books.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.
- Hume, D. (1938). *An abstract of a treatise of human nature*. London: Cambridge University Press. (Original work published 1740)
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136–153.
- Kahneman, D., & Tversky, A. (1982a). 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.
- Kahneman, D., & Tversky, A. (1982b). Variants of uncertainty. *Cognition*, *11*, 143–157.
- Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology*, *59*, 1101–1110.
- Kao, S.-F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1363–1386.
- Keil, F. C., & Wilson, R. A. (Eds.). (2000). *Explanation and cognition*. Cambridge, MA: Cambridge University Press.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192–238). Lincoln: University of Nebraska Press.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, *28*, 107–128.
- Lewis, N. A. (2002, May 30). F.B.I. chief admits 9/11 might have been detectable. *The New York Times*. Retrieved May 30, 2002, from <http://www.nytimes.com>
- Lipe, M. G. (1991). Counterfactual reasoning as a framework for attribution theories. *Psychological Bulletin*, *109*, 456–471.
- Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford, England: Oxford University Press.
- Mandel, D. R. (2003a). Counterfactuals, emotions, and context. *Cognition & Emotion*, *17*, 139–159.
- Mandel, D. R. (2003b). Effect of counterfactual and factual thinking on causal judgments. *Thinking & Reasoning*, *9*, 245–265.
- Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology*, *71*, 450–463.
- Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General*, *127*, 269–285.
- Mandel, D. R., N'gbala, A., & Bonnefon, J.-F. (2001). [Tests of facilitation in causal and counterfactual judgment]. Unpublished raw data.
- Markman, K. D., Gavanski, I., Sherman, S. J., & McMullen, M. N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology*, *29*, 87–109.
- McCloy, R., & Byrne, R. M. J. (2000). Counterfactual thinking about controllable events. *Memory & Cognition*, *28*, 1071–1078.
- McClure, J., Hilton, D. J., Cowan, J., Ishida, L., & Wilson, M. (2001). When people explain difficult actions, is the causal question how or why? *Journal of Language and Social Psychology*, *20*, 339–357.
- McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology*, *57*, 189–200.
- McGill, A. L. (1998). Relative use of necessity and sufficiency information in causal judgments about natural categories. *Journal of Personality and Social Psychology*, *75*, 70–81.
- McGill, A. L., & Klein, J. G. (1993). Contrastive and counterfactual

- thinking in causal judgment. *Journal of Personality and Social Psychology*, 64, 897–905.
- McGill, A. L., & Tenbrunsel, A. E. (2000). Mutability and propensity in causal selection. *Journal of Personality and Social Psychology*, 79, 677–689.
- Michotte, A. E. (1963). *The perception of causality* (T. R. Miles & E. Miles, Trans.). New York: Basic Books. (Original work published 1946)
- Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology*, 59, 1111–1118.
- N'gbala, A., & Branscombe, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology*, 31, 139–162.
- Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, 52, 111–139.
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, 62, 189–206.
- Piaget, J. (1974). *Understanding causality*. New York: Norton.
- Platt, J. R. (1964, October 16). Strong inference. *Science*, 146, 347–353.
- Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology*, 52, 288–302.
- Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, 130, 323–360.
- Roese, N. J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, 66, 805–818.
- Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121, 133–148.
- Roese, N. J., Hur, T., & Pennington, G. L. (1999). Counterfactual thinking and regulatory focus: Implications for action versus inaction and sufficiency versus necessity. *Journal of Personality and Social Psychology*, 77, 1109–1120.
- Roese, N. J., & Olson, J. M. (1997). Counterfactual thinking: The intersection of affect and function. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 1–59). San Diego, CA: Academic Press.
- Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, 110, 101–120.
- Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.
- Shultz, T. R., Fisher, G. W., Pratt, C. C., & Rulf, S. (1986). Selection of causal rules. *Child Development*, 57, 143–152.
- Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, 126, 323–348.
- Spellman, B. A., & Kincannon, A. (2001). The relation between counterfactual (“but for”) and causal reasoning: Experimental findings and implications for jurors’ decisions. *Law and Contemporary Problems*, 64, 241–264.
- Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. *Current Directions in Psychological Science*, 8, 120–123.
- Spellman, B. A., & Mandel, D. R. (2003). Causal reasoning, psychology of. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol. 1, pp. 461–466). London: Nature.
- Sperber, D., Premack, D., & Premack, A. J. (Eds.). (1995). *Causal cognition: A multidisciplinary debate*. Oxford, England: Clarendon Press.
- Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam, the Netherlands: North-Holland.
- Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization–minimization hypothesis. *Psychological Bulletin*, 110, 67–85.
- Tetlock, P. E., & Belkin, A. (1996). Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives. In P. E. Tetlock & A. Belkin (Eds.), *Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives* (pp. 3–38). Princeton, NJ: Princeton University Press.
- Vinokur, A., & Ajzen, I. (1982). Relative importance of prior and immediate events: A causal primacy effect. *Journal of Personality and Social Psychology*, 56, 161–169.
- Weiner, B. (1985). “Spontaneous” causal thinking. *Psychological Bulletin*, 109, 74–84.
- Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, 56, 161–169.
- Wells, G. L., Taylor, B. R., & Turtle, J. W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53, 421–430.

(Appendixes follow)

Appendix A

Vignette Used in Experiment 1

Mr. Wallace is highly influential in the organized crime scene. Although he commands the respect of many, there are also a number of people who stand to benefit from his death. Today, he was having lunch with one of his associates, Mr. Vincent. At one point during their meeting, Mr. Wallace left the table to go to the toilet. While he was gone, Mr. Vincent managed to slip a lethal dose of poison into his martini. The poison is known to take one hour (give or take 10 minutes, depending on the victim's body weight) to kill and would go unnoticed for the first half hour after consumption. When Mr. Wallace returned to the table, he finished his lunch and then drank the rest of his martini. The two men concluded their meeting a few minutes later. Mr. Wallace paid the bill and they both left the restaurant going in different directions.

Mr. Wallace had another important business meeting about 15 minutes away in a remote part of the countryside. On an isolated stretch of road, he noticed that he was suddenly being followed by a van that had pulled out from the roadside. The van accelerated and began hitting his car from behind. Mr. Wallace recognized that the driver was Mr. Bruce, an associate of one of Mr. Wallace's rivals. He tried to evade the van but there was no time. The van relentlessly hit his car and pushed it off the side where it fell into a ravine and exploded. The remains of Mr. Wallace's body were discovered later that day. The chief coroner's report later revealed that Mr. Wallace had received fatal burns in the car explosion. The report also indicated, however, that a lethal dose of poison was found in Mr. Wallace's blood.

Appendix B

Vignette Used in Experiment 2

Long ago, when John was only 17 years old, he got a job working for a large manufacturing company. He started out working on an assembly line for minimum wage, but after a few years at the company, he was given a choice between two line manager positions. He could stay in the woodwork division, which is where he was currently working. Or he could move to the plastics division. John was unsure what to do because he liked working in the woodwork division, but he also thought it might be worth trying something different. He finally decided to switch to the plastics division and try something new.

For the last 30 years, John has worked as a production line supervisor in the plastics division. After the first year there, the plastics division was moved to a different building with more space. Unfortunately, through the many years he worked there, John was exposed to asbestos, a highly carcinogenic substance. Most of the plastics division was quite safe, but the small part in which John worked was exposed to asbestos fibers. And now,

although John has never smoked a cigarette in his life and otherwise lives a healthy lifestyle, he has a highly progressed and incurable case of lung cancer at the age of 50. John had seen three cancer specialists, all of whom confirmed the worst: that, except for pain, John's cancer was untreatable and he was absolutely certain to die from it very soon (the doctors estimated no more than 2 months).

Yesterday, while John was in the hospital for a routine medical appointment, a new nurse accidentally administered the wrong medication to him. John was allergic to the drug and he immediately went into shock and experienced cardiac arrest (a heart attack). Doctors attempted to resuscitate him but he died minutes after the medication was administered.

Received July 31, 2001

Revision received March 31, 2003

Accepted April 1, 2003 ■

Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at journals@apa.org and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.