

Mandel, D. R. (2003). Effect of counterfactual and factual thinking on causal judgments. *Thinking & Reasoning*, 9, 245-265.

Effect of Counterfactual and Factual Thinking on Causal Judgments

David R. Mandel
University of Victoria

The significance of counterfactual thinking in the causal judgment process has been emphasized for nearly two decades, yet no previous research has directly compared the relative effect of thinking counterfactually versus factually on causal judgment. Three experiments examined this comparison by manipulating the task frame used to focus participants' thinking about a target event. Prior to making judgments about causality, preventability, blame, and control, participants were directed to think about a target actor either in counterfactual terms (what the actor could have done to change the outcome) or in factual terms (what the actor had done that led to the outcome). In each experiment, the effect of counterfactual thinking did not differ reliably from the effect of factual thinking on causal judgment. Implications for research on causal judgment and mental representation are discussed.

Since Kahneman and Tversky's (1982) seminal paper on the simulation heuristic, many studies have examined how comparing reality to other possible worlds, or thinking counterfactually, can influence people's reactions to events (for a review, see Roese, 1997). An important theoretical claim in this body of literature has been that counterfactual thinking can influence the causal judgment process in two, distinct ways. The first proposed route of influence involves the notion of contrast between factual and counterfactual events. When an outcome evokes surprise by violating expectancies or norms, the persistence of attention to those expectancies is, by definition, a counterfactual representation because, in hindsight, individuals know that their violated expectancies are not factual representations of the past. Kahneman and Miller (1986) proposed that this form of counterfactual thinking, in which an expectancy-consistent counterfactual outcome contrasts sharply with a surprising factual outcome, can influence the causal judgment process by prompting the question, Why? Similarly, Hilton and Slugoski (1986) and McGill (1989), building on the work of earlier philosophers (Gorovitz, 1965; Hart & Honoré, 1959; Hesslow, 1983), proposed that the normal outcomes evoked by counterfactuals can serve as contrast cases that define the focus of explanation.

The second proposed route of influence involves the notion that counterfactual conditionals are well-suited for spotlighting causal contingencies. That is, some theorists (e.g., Roese & Olson, 1995a) have argued that the conditional structure of counterfactuals that take an "if only..., then..." form serve an important function in the causal selection process. For example, after a student learns that she just failed an important exam, she might think something like "If only I had studied a bit harder, I would have passed the exam." Although these counterfactuals may simply reflect existing causal beliefs, their activation may also strengthen or reaffirm causal beliefs, perhaps by making explicit knowledge of contingencies and "but for" causes that would otherwise have received little weight in subsequent thought and judgment (see Tetlock & Belkin, 1996). A strong version of this argument holds that counterfactual conditionals represent tests of hypothesized causal connections between the outcome and antecedent events in an episode. As Mackie (1974) put it, "... in quite a primitive and unsophisticated way we can transfer the non-occurrence of Y from the before

situation to a supposed later situation, in which, similarly, X did not occur, and form the thought which is expressed by the statement ‘If X had not occurred, Y would not have occurred’” (p. 56).

Psychological accounts of causal judgment that build on Mackie (1974) posit that counterfactual conditionals can act as tests of the necessary (or “but for”) causes of outcomes that demand explanation (e.g., Einhorn & Hogarth, 1986; Kahneman & Tversky, 1982; Lipe, 1991; McGill & Klein, 1993; Roese & Olson, 1997; Wells & Gavanski, 1989). Although no study to my knowledge has demonstrated that people intentionally generate counterfactual conditionals in order to test their hypotheses about causal relations, there is some empirical support for the weaker notion that causal judgments are influenced by the ease with which people can mentally undo focal events in the past and by the content of their counterfactual musings (Branscombe, Owen, Garstka, & Coleman, 1996; Roese & Olson, 1997; Wells & Gavanski, 1989). The case for an effect of counterfactual thinking on causal judgment is strengthened by studies that have also examined related judgments of blame (Branscombe et al., 1996; Miller & Gunasegaram, 1990) and controllability (Roese & Olson, 1995b). Taken together, however, this body of evidence is based almost exclusively on either correlational studies or experiments in which the mutability of focal outcomes is manipulated (i.e., manipulations of the ease with which alternative outcomes can be mentally constructed) (for an exception to this rule, see Roese & Olson, 1997, who report an experiment using a task facilitation paradigm). In experiments that manipulate outcome mutability, it is often assumed that if causal judgments in the high-mutability condition are more extreme or frequent than in comparable low-mutability conditions, then counterfactual thinking must have played a causal role (e.g., Wells & Gavanski, 1989).

Other research (e.g., N’gbala & Branscombe, 1995), however, indicates that the effect of mutability manipulations on causal judgments may have had more to do with the specific hypothetical scenarios used in previous experiments than with a robust effect of counterfactual thinking on causal selection. For instance, Mandel and Lehman (1996, Experiment 3) showed that, when participants read a case that afforded the opportunity to select different antecedents in counterfactual and causal judgment tasks, mutability manipulations influenced counterfactual assessments but had no reliable effect on causal judgments.

Importantly, no research to date has directly addressed the basic question of whether the effect of counterfactual thinking on attribution differs in reliable and consequential ways from the effect of factual thinking. That is, does thinking counterfactually about what might, could, or would have been have a greater impact on attribution than thinking factually about what was? The closest study in this regard is one by McCloy and Byrne (2002, Experiment 1). Participants in one condition were instructed to think counterfactually about a hypothetical case that they had just considered by completing an “if only...” sentence stem, whereas participants in a control condition were not asked to do so. Participants then rated the importance of a single causal candidate. Mean causal importance ratings did not differ between the counterfactual and control conditions. As McCloy and Byrne (2002) acknowledged, however, the null finding in their study is difficult to interpret because the completion task did not constrain attention to a specific antecedent, whereas the causal rating task did.

Given the strong claims that have been made about the consequential effects of counterfactual thinking on causal judgment and related attribution processes, research that

directly compares the effect of counterfactual and factual thinking on dependent measures such as causality, prevention, blame, and control is sorely needed. For instance, consider Roese and Olson's (1995a) claim that although "...not all conditionals are causal...counterfactuals, by virtue of the falsity of their antecedents, represent one class of conditional propositions that are always causal. The reason for this is that with its assertion of a false antecedent, the counterfactual sets up an inherent relation to a factual state of affairs" (p. 11). This proposal suggests not only that counterfactual conditional thinking can influence causal judgment, but that it should have a stronger effect on causal judgment than comparable factual conditional thinking, which would not necessarily automatically recruit a contrast case. Recent extensions of mental model theory (e.g., Byrne, 1997; Byrne & Tasso, 1999) have reached similar conclusions—namely that, whereas counterfactual conditionals automatically “recover” factual models, thus establishing a salient contrast case, factual conditionals do not automatically recover counterfactual models. These accounts do not hedge on the prediction that counterfactual and factual modes of thinking will produce different effects on causal attribution and prompt different cognitive representations.

In contrast with theoretical accounts that accord special status to counterfactual thinking in the causal judgment process, the present account posits that counterfactual and factual statements often may be viewed as (quasi-)complementary descriptions, or “frames,” that, roughly speaking, mean the same thing. For example, the statement, “If only I hadn't left work late, I would have avoided the accident,” can be loosely reframed as, “Other things being equal, given that I left work late, I was caught in the accident.” The former statement takes a counterfactual form and the latter statement takes a factual form but, importantly, the gist of the statement is preserved. The present account coheres with Lebow's (2000) recent analysis of the role of counterfactual and factual analysis in the interpretation of world politics. Lebow states that “[g]ood counterfactual thought experiments differ little from ‘factual’ modes of historical reconstruction” (p. 555), and adds that “[a]ny sharp distinction between factials and counterfactuals rests on questionable ontological claims” (p. 556).

The present account further posits that the manner in which people frame the content of their thinking—counterfactually or factually, in this case—will have a relatively small effect on causal judgment because the shift from one frame to another involves primarily a change of structure but not a change of construed meaning. In this sense, the reframing of counterfactuals into factials and vice versa differs from the reframing of prospects in terms of either gains or losses (Tversky & Kahneman, 1981; but see also Mandel, 2001, for an analysis of methodological and theoretical limitations of research in the latter domain). Correspondingly, it was predicted that the manner in which a thinking directive is framed—namely, to think about what might have been as opposed to what was—would have, at best, a weak effect on causal and related judgments.

Research Overview

The present research directly examined the relative effect of counterfactual versus factual thinking focus on causal judgment (broadly construed to include the focal measures). In three experiments, participants were asked to bring to mind either a counterfactual thought about how the target outcome could have turned out differently than it did or a factual thought about how the target outcome turned out as it did. Following the thinking-focus manipulation, participants indicated their level of agreement with four statements about causality, preventability, blame, and control. Experiment 1 examined the effect of thinking focus on judgments about recollected negative interpersonal conflicts. Participants were

asked to recall an interpersonal experience and, prior to the judgment task, to generate a thought either about what they (or someone else) might have done to improve the outcome (counterfactual condition) or about what they (or someone else) had in fact done that led events to turn out as they did (factual condition). Experiment 2 built on the previous experiment (a) by including a control group in which no directive to think either counterfactually or factually was provided and (b) by generalizing the research to positive interpersonal experiences. Experiment 3 extended the research by independently manipulating thinking focus and action focus: participants were asked to focus on either a counterfactual/factual omission (e.g., “if only Jones had not done X...” or “given that Jones didn’t do X...””) or a counterfactual/factual commission (e.g., “if only Jones had done X...” or “given that Jones did X...”).

Experiment 1

The key objective of Experiment 1 was to examine the effect of thinking focus on participants’ causal judgments about an interpersonal conflict that they had experienced. A second objective was to examine whether the length of time since the conflict would influence the effect of thinking focus on causal judgment. As time elapses following an event, the understanding that an individual arrives at by thinking counterfactually and factually are likely to be integrated in memory. Consequently, the effect of thinking focus on causal judgment was predicted to decrease as a function of time since the event.

Method

One hundred and ninety-two undergraduates from University of Hertfordshire (Hatfield, England) participated in the experiment on a voluntary basis. Participants were randomly assigned to one of four conditions in a 2 (Person Focus: self, other) × 2 (Thinking Focus: counterfactual, factual) between-participants design. At the start of the experiment, participants were asked to think of an interpersonal conflict that has been significant to them and to briefly describe it, indicating also how long ago the conflict began. Participants in the self condition were asked to think about either something they “could have done differently that would have led things to turn out better than they actually did” (the counterfactual condition) or about something they “had done that led things to turn out as they actually did” (the factual condition). Participants in the other condition were asked to think about something that another person involved in the conflict either could have done differently to improve the outcome or actually did that led to the conflict.

Following the thinking-focus manipulation, participants indicated their level of agreement with a series of statements that ascribed causality, (foregone) preventability, blame, and control to the target actor whose role participants had just been asked to consider. That is, participants in the self condition judged statement about themselves, whereas participants in the other condition judged statements about the person that they had identified in response to the preceding thinking directive. For instance, in the self condition, participants were asked to judge the following statements: (a) “I personally caused the negative event to happen,” (b) “I personally could have prevented the negative event,” (c) “I am to blame for the negative event,” and (d) “I had control over the outcome of the event.” In the other condition, the subject of these statements was changed to “He/She.” Participants indicated their responses to the statements on a 7-point Likert scale that ranged from strongly disagree (–3) to strongly agree (+3). The order of the first two statements (viz., causality and preventability) was counterbalanced.

Results and Discussion

Participants' interpersonal conflicts fell into three content categories: 40% of participants recalled a conflict with a friend, 30% recalled a family conflict, and 30% recalled a conflict with an intimate partner. The amount of time since the event did not differ as a function of either person focus or thinking focus, and the mean was 16.06 months ($SD = 33.13$). Nor did time since the conflict correlate significantly (two-tailed $\alpha = .05$) with any of the dependent measures. The correlations among the dependent measures, as well as their mean values and standard deviations, are shown in Table 1. Given the positive correlations among these measures, the four ratings were averaged to provide a single causal judgment scale (Cronbach's $\alpha = .72$).

The sample was divided into three groups on the basis of reported time since the conflict. The recent group ($n = 69$) experienced the conflict up to and including 1 month ago, the intermediate group ($n = 60$) experienced the conflict from 1.5 to 9 months ago, and the distant group ($n = 63$) experienced the conflict from 10 months to 23 years ago. A 2 (Person Focus) \times 2 (Thinking Focus) \times 3 (Time) between-participants ANOVA on the scale values revealed a significant main effect of person focus ($F[1, 180] = 23.39$, $MSE = 2.47$, $p < .001$) and time ($F[2, 180] = 3.12$, $p = .047$), and a thinking focus \times time interaction effect ($F[2, 180] = 4.01$, $p < .020$). No other effect was significant (test-wise, two-tailed $\alpha = .05$). Participants were more likely to agree with the causal statements when they implicated the other individual ($M = 0.71$) than when they implicated themselves ($M = -0.29$). Additional analyses showed that person focus did not interact with type of event (friend vs. family. vs. sexual partner), $F < 1$. The main effect of person focus is consistent with research on the self-serving bias (e.g. Greenberg, Pyszczynski, & Solomon, 1982; Miller & Ross, 1975), which indicates that people tend to attribute successes to their own actions and failures to the actions of others or to other external factors. Importantly, however, this effect was not influenced by whether participants were directed to think counterfactually or factually about themselves or another person involved in the conflict.

The main effect of time was qualified by the predicted interaction with thinking focus. As Table 2 shows, participants in the recent group were more likely to agree with the causal statements focusing on the target actor if they first thought factually rather than counterfactually about that actor's role, $t(67) = 3.14$, $p = .003$. In contrast, the pattern was reversed in the intermediate and distant groups, although the effect of thinking focus was not significant at either time level ($ps > .25$). Thus, the prediction that the effect of thinking focus would be inversely related to time since the target incident was supported. Contrary to the popular notion that counterfactual thinking plays a special role in shaping causal judgments, however, factual thinking had a greater impact on judgments than counterfactual thinking among those participants who recalled an event that had happened in the relatively recent past.

One possible explanation for the null main effect of thinking focus is that the manipulation had different effects on the four individual judgments, which cancelled out in the averaging process used to create the judgment scale. If that were the case, we should find a significant interaction effect of question and thinking focus. A 2 (Thinking Focus) \times 4 (Question [repeated measure]) mixed ANOVA, however, revealed that the interaction was unreliable, $F(3, 188) < 1$. Therefore, the null main effect of thinking focus cannot be attributed to a cancellation of reliable effects across questions. Nor can it be due to a lack of statistical power given the large sample size in this experiment.

Experiment 2

A limitation of Experiment 1 was that it did not include a control condition in which participants were neither asked to think counterfactual nor factually about the target event prior to judgment. It is possible, therefore, that the manipulation was merely ineffective. The key objective of Experiment 2 was to examine whether either thinking focus would yield an effect on causal judgments that differed significantly from a control group not receiving any thinking directive. Once again, participants were asked to recall an interpersonal event and rate their level of agreement with statements that attributed causality, preventability, blame, and control to either themselves or another person involved in the recalled event. In Experiment 2, however, participants were asked to think of a significant event that occurred within the last three months or as close to that time frame as possible in order to restrict the range of time since the recalled event to a relatively recent period. Moreover, in an attempt to increase the generalizability of the results, the valence of the recalled event was manipulated. Half the sample was asked to recall a negative interpersonal experience, as in Experiment 1, whereas the other half was asked to recall a positive interpersonal experience.

If the thinking-focus manipulation is effective, then a reasonable prediction is that the mean causal judgment in the control condition would be less positive (indicating weaker attribution to the target) than in the experimental (i.e., counterfactual and factual) conditions. A more sophisticated prediction, however, is that the expected effect of thinking focus just noted would be manifested more strongly when participants were judging the influence of the other person rather than themselves. The rationale for this predicted interaction between thinking focus and person focus is that people tend to think much more frequently and in greater depth about the effects of their own behavior on important events in their lives than about the effects of other people's behavior on these events (e.g., Miller & Taylor, 1995). For example, Mandel and Lehman (1996, Experiment 1) found that participants who read a scenario about a car accident tended to undo the behavior of the actor whose perspective they were instructed to adopt (see also Kahneman & Tversky's, 1982, discussion of focus effects). Consequently, it was predicted that, whereas causal knowledge in the self condition should be fairly automatic by three months after the event, and the results of counterfactual and factual thinking well-integrated, knowledge regarding others should be relatively less articulated and more susceptible to thinking-focus manipulations.

Method

Ninety-six undergraduates from University of Hertfordshire participated in the experiment on a voluntary basis. Participants were randomly assigned in equal numbers to one of 12 conditions in a 2 (Person Focus: self, other) \times 3 (Thinking Focus: counterfactual, factual, control) \times 2 (Event Valence: positive, negative) between-participants design. Participants were asked to recall an interpersonal event that has been significant to them and that had occurred within the last three months, and to indicate how long ago the event took place. Then, participants were instructed to take a few moments to think vividly about the experience and about how they felt at the time.

In the negative-event condition, as in Experiment 1, participants were asked to think of a significant interpersonal conflict. Participants in the positive-event condition were asked to "think of a positive interpersonal experience that has been significant to you." Next, person focus and thinking focus were manipulated. The person-focus manipulation was identical to that in Experiment 1, directing attention to either the participant or another person involved in the event. In the negative-event condition, the counterfactual- and factual-thinking prompts were worded as in Experiment 1. In the positive-event condition, participants (in the self

condition) were asked to think about either something they "could have done differently that would have led things to turn out worse than they actually did" (the counterfactual condition) or something they "had done that led things to turn out as they actually did" (the factual condition). The control group was not asked to think about either a counterfactual or a factual thought. The dependent measures were identical to those used in Experiment 1, with one exception. In the positive-event condition, the preventability question was changed to read "I (He/she) personally could have altered the positive event" because people typically do not try to prevent positive outcomes. The order of the causality and preventability questions was counterbalanced as in Experiment 1.

Results and Discussion

The mean reported time since the recalled event was 3.01 months ($SD = 5.91$). The correlations among the dependent measures, as well as their mean values and standard deviations, are shown in Table 1. Each correlation was significant, and the magnitudes of the correlations had the same rank ordering as in Experiment 1. Once again, the four ratings were averaged to provide a single, causal judgment scale (Cronbach's $\alpha = .72$).

A 2 (Person Focus) \times 3 (Thinking Focus) \times 2 (Event Valence) between-participants ANOVA on the causal judgment ratings was conducted. As predicted, this analysis revealed a significant thinking focus \times person focus interaction effect, $F(2, 84) = 5.70$, $MSE = 1.28$, $p = .005$. No other effect was significant (test-wise, two-tailed $\alpha = .05$). Table 3 shows the mean ratings as a function of thinking focus and person focus. Simple-effects tests revealed that, as predicted, mean causal judgment in the self condition did not differ significantly as a function of thinking focus, $F(2, 45) < 1$. In contrast, as anticipated, mean causal judgment in the other condition differed significantly as a function of thinking focus, $F(2, 45) = 9.13$, $p < .001$. Pair-wise comparisons using Tukey HSD tests confirmed that the mean rating in the control condition was significantly lower than in the factual condition ($p = .004$) and in the counterfactual condition ($p = .001$). Importantly, however, the mean ratings in the latter two conditions did not differ significantly ($p > .85$). Moreover, as in Experiment 1, a 3 (Thinking Focus) \times 4 (Question [repeated measure]) mixed ANOVA revealed that the interaction effect was unreliable, $F(6, 184) = 1.25$, $p > .28$. Therefore, the null main effect of thinking focus, once again, cannot be attributed to a cancellation of different effects across judgments.

Consistent with Experiment 1, the findings of Experiment 2 do not support the view that counterfactual thinking influences causal judgment in a manner that reliably differs from factual thinking. Importantly, however, unlike Experiment 1, Experiment 2 demonstrated that, compared to a baseline measure, an explicit thinking focus yielded the predicted effect on causal judgments. Taken together, the results of Experiments 1 and 2 support the notion that factual and counterfactual thinking can yield similar effects on causal judgment because the transformation of structure does not have a significant impact on meaning.

Experiment 3

In Experiments 1 and 2, the wording of the counterfactual prompt to think of something "you could have done differently that would have led things to turn out better than they actually did" may have been interpreted narrowly by participants as a request for additive counterfactuals, in which a factual omission is mentally undone, to be brought to mind (cf. Roese & Olson, 1993). In contrast, the wording of the factual prompt to think of something "you had done that led things to turn out as they actually did" may have been interpreted narrowly by participants as a request for factual commissions to be brought to

mind. Experiment 3 was designed to independently manipulate these factors. Moreover, a related interest involved examining whether the agency effect—namely, the tendency for people to judge that actions leading to a disappointing outcome will make an actor feel worse than inactions leading to the same outcome (e.g., Byrne & McEleney, 2000; Kahneman & Tversky, 1982; N’gbala & Branscombe, 1997)—generalizes to causal judgments. If the agency effect does generalize, then participants should tend to agree more strongly with statements that implicate an actor when they had just thought about something the actor did rather than something he didn’t do that led to the negative outcome, or when they had just thought about something the actor could have avoided doing (thus mentally negating an action) rather than something he could have done (thus mentally negating an inaction) that might have undone the negative outcome.

Finally, to rule out the possibility that prior thinking about the event eliminated the effect of thinking focus in Experiments 1 and 2, Experiment 3 used a hypothetical vignette of a negative outcome as a stimulus event. Thus, if there is an important difference between counterfactual and factual thinking, it should be most easily detected before the two forms can become integrated in knowledge about the target event. Alternatively, if no reliable main effect of thinking focus is observed even after considering a novel episode, then support for the “integration of knowledge” explanation would be considerably weakened.

Method

One hundred and sixty undergraduates from University of Victoria (Victoria, B.C., Canada) participated in the experiment for course credit. Participants were randomly assigned in equal numbers to one of four conditions in a 2 (Thinking Focus: counterfactual, factual) × 2 (Action Focus: omission, commission) between-participants design.

Participants first read the following vignette (cf. Kahneman & Tversky, 1982; Mandel & Lehman, 1996):

Mr. Jones is 47 years old, the father of three and a successful banking executive. His wife has been ill at home for several months. On the day of his accident, Mr. Jones left his office at the regular time. He occasionally left early to take care of home chores at his wife's request, but this was not necessary on that day. Mr. Jones did not drive home by his regular route. The day was exceptionally clear so Mr. Jones decided to drive along the shore to enjoy the view.

The accident occurred at a major intersection. The light turned yellow as Mr. Jones approached. Witnesses noted that he braked hard to stop at the crossing, although he could easily have gone through. His family recognized this as a common occurrence in Mr. Jones' driving. As Mr. Jones began to cross after the light changed, a car charged into the intersection at high speed and rammed his car from the left. Mr. Jones was seriously injured. The other driver, an elderly man, had lost control after suffering a mild stroke while he was driving.

Next, participants were instructed to bring to mind a particular type of thought about the accident. Specifically, participants were either asked to “think for a moment about something that Mr. Jones”:

- a. “could have done that would have kept him from being in the accident” (counterfactual commission).
- b. “could have avoided doing that would have kept him from being in the accident” (counterfactual omission).

- c. “did that led him to be in the accident” (factual commission).
- d. “didn’t do that led him to be in the accident” (factual omission).

In an attempt to strengthen the manipulations, all participants were further instructed to “focus intensely for a few moments on the thought that you just brought to mind. Repeat it in your mind for at least 30 seconds.” Next, participants completed the rating task, which was of the same form as in the previous experiments. In Experiment 3, the statements always focused on the protagonist (Mr. Jones) as the target of judgment. For instance, for the causal judgment question, participants rated the extent to which they agreed with the statement “Mr. Jones caused the accident.” The order of the causality and preventability items, once again, was counterbalanced across participants.

Results and Discussion

As in the first two experiments, the four items were averaged to provide an overall causal judgment measure (Cronbach’s $\alpha = .74$). Table 4 shows the mean judgments as a function of thinking focus and action focus. A 2 (Thinking Focus) \times 2 (Action Focus) between-participants ANOVA on causal judgment scale scores revealed no significant main effect of thinking focus ($F[1, 156] = 1.14$, $MSE = 1.59$, $p = .29$) or action focus ($F[1, 156] = 2.36$, $p > .12$). Nor was the interaction effect reliable ($F < 1$). Moreover, as in Experiments 1 and 2, the 2 (Thinking Focus) \times 4 (Question) interaction effect was unreliable, $F(3, 156) < 1$. Therefore, despite the fact that Experiment 3 independently manipulated thinking focus and action focus and strengthened the former manipulation by having participants intensely focus on their thoughts prior to the judgment task, the findings did not support the notion that counterfactual and factual thinking differentially influence causal judgment. Importantly, by using a scenario to which participants had no previous exposure, the null main effect of thinking focus could neither be explained in terms of the “integration of knowledge” hypothesis discussed earlier, nor in terms of a cancellation of reliable effects across judgments.

To test for evidence of an agency effect on causal judgment, the counterfactual-omission (i.e., negate an action) and factual-commission (i.e., take an action) conditions were combined into an action condition and the counterfactual-commission (i.e., negate an inaction) and factual-omission (i.e., do not act) conditions were combined into an inaction condition. Mean judgments following a directive to focus on the protagonist’s action ($M = -0.98$) did not differ significantly from mean judgments following a directive to focus on the protagonist’s inaction ($M = -0.80$), $t(158) < 1$. That is, no reliable agency effect on causal judgment was observed.

General Discussion

The primary objective of the present research was to examine whether judgments of causality, preventability, blame, and control would differ depending on whether participants’ attention was focused on either what they might have done to change the event (counterfactual thinking) or what they actually did to make it turn out the way it did (factual thinking). Without exception, the findings of the three experiments reported here do not indicate that counterfactual thinking has a stronger influence on causal judgment than factual thinking. Indeed, in Experiment 1, a factual-thinking focus had a stronger effect on mean judgment than a counterfactual-thinking focus among participants who experienced the conflict within one month of the experiment. Experiment 2 further demonstrated that prompts to think either counterfactually or factually intensified causal attribution to the focal target of participants’ thoughts compared to a control condition in which neither form of thinking was

explicitly prompted. Taken together, the present findings suggest that counterfactual and factual thinking have similar effects on causal judgment, contrary to contemporary accounts of the causal judgment process that have emphasized the special role of counterfactual thinking (e.g., Branscombe et al., 1996; Lipe, 1991; Roese & Olson, 1995a; Roese, 1997; Wells & Gavanski, 1989).

The present findings cohere strongly with recent proposals concerning the overestimated difference between factual and counterfactual analyses in causal explanation. As Lebow (2000) noted, “even when evidence is meagre or absent, the difference between counterfactual and ‘factual’ history may still be marginal. Documents are rarely ‘smoking guns’ that allow researchers to establish motives or causes beyond a reasonable doubt” (p. 553). In everyday life too, factual representations can often be easily reframed as counterfactual representations with little change in meaning, just as omissions can often be easily reframed as commissions (Davis, Lehman, Wortman, Silver, & Thompson, 1995). Consider, for example, a person who says, “I barely caught the train.” This factual statement differs trivially in meaning from the counterfactual statement, “I almost missed the train.” Both statements, when semantically unpacked, highlight the “fact” that the train was caught in time but that this result was the outcome of a previously uncertain state in which the propensity for not catching the train was high. More generally, the psychology of close-counterfactual “almost” thinking, which has received some attention (Kahneman & Varey, 1990; Tetlock, 1998), seems trivially different from the psychology of close-factual “barely” thinking, which has received none.

Nor does the line between factual and counterfactual claims blur only in cases involving ostensibly close possible worlds. Consider a case in which there is a strong dispositional probability of Smith suffering a stroke. Once again, the factual statement that “given Smith’s high cholesterol level, it’s not surprising that he suffered a stroke” appears to differ trivially in meaning from the counterfactual statement, “if only Smith didn’t have high cholesterol, he might not have suffered the stroke.” The two statements reframe the same quantum of causal knowledge (and both statements hedge somewhat relative to the explicit causal statement that “Smith had a stroke because he had high cholesterol”).

Implications for Theories of Causal Judgment

The idea that causal judgment critically depends on counterfactual reasoning assumes that causal judgment is geared primarily toward identifying the necessary (“but for”) causes of a target outcome: But for X, Y would not have occurred (Lipe, 1991; Mackie, 1974). Unqualified, this idea is indefensible. First, as the legal philosophers Hart and Honoré (1959) noted long ago, in order for necessary, “but for” conditions to be deemed causally relevant they must appeal to our commonsense notions of causality. An example from Spellman (1997) illustrates the point: “If a motorist speeds for a while, then slows down, then minutes later hits a child crossing the street, the speeding is not considered a true cause of the accident even though the child would have crossed safely had the motorist not sped” (p. 324). Thus, meeting the criterion of conditional necessity is not sufficient for ascribing causal status to an antecedent.

Another problem is that counterfactual conditional thinking can only identify the necessary generative causes of an outcome, not the sufficient generative causes. In contrast, several lines of research now indicate that greater weight is given to a sufficiency criterion than a necessity criterion in causal judgment, especially in reasoning causally about natural categories (McGill, 1998). For instance, Mandel and Lehman (1998, Experiment 1) found

that when participants were asked to define causality in their own terms, they almost invariably provide definitions in terms of sufficiency, such as “if the cause is present, the effect will occur,” whereas only a small minority additionally define causality in terms of necessity (e.g., “if the cause is absent the effect won’t occur”). This preference for defining causality in terms of the sufficiency of the causal candidate to generate the effect has also been replicated using a forced choice response format (Mandel, 2002, Experiment 2). Moreover, the greater weight given to contingency information from the A-cell (cause-present and effect-present) and B-cell (cause-present and effect-absent) than from the C-cell (cause-absent and effect-present) and D-cell (cause-absent and effect-absent) in studies of causal induction (e.g., Mandel & Lehman, 1998; Schustack & Sternberg, 1981; Wasserman, Dorner, & Kao, 1990) indicates that people think about causality primarily in terms of a “cause-present” sufficiency criterion than in terms of a “cause-absent,” necessity criterion (for an analysis of the relation between the weighting of causal judgment criteria and contingency information, see Mandel & Lehman, 1998).

Finally, recent research (see Mandel, 2002, Spellman & Kincannon, 2001) demonstrates that in cases in which there are multiple sufficient causes, participants’ counterfactual judgments either tend to undo all of the sufficient causes simultaneously or they tend to undo a single necessary precondition for the set of sufficient causes. In contrast, participants tend to ascribe a high degree of importance to each sufficient cause on its own in their causal judgments. For example, Mandel (2002, Experiment 1) presented participants with a case in which there were two sufficient causes of the protagonist’s death: a poisoning episode followed by an independent car-crash episode. The car crash was viewed as the primary cause, followed by the poisoning, which, although sufficient to cause death did not directly contribute to the manner in which the protagonist died. In contrast, when asked to undo the protagonist’s death, the modal response focused on negating his involvement in a life of organized crime. This factor, however, was not viewed as a good causal explanation of the protagonist’s death. Such dissociations in judgment focus do not support the argument that counterfactual thinking (or even outcome mutability) is a requisite condition for causal judgment.

Implications for Mental Representation Theories

According to recent work on mental model theory (e.g., Byrne, 1997; Byrne & McEleney, 2000; Byrne & Tasso, 1999), counterfactual representations are more fleshed out than factual representations because, whereas mental models of factual events constitute default representations, as Byrne (1997) stated, “counterfactual scenarios need to be recoverable from the factual scenario on which they are based” (p. 107). This proposed asymmetry in model recoverability—namely, that a counterfactual model automatically recovers the factual model but not vice versa—provides a cognitive explanation for why counterfactuals are good at spotlighting contrast cases that orient causal reasoning. For without a theory that posits such an asymmetry at the representational level, one should wonder why factual descriptions are not equally effective at setting up contrast cases by automatically recovering a counterfactual model.

The recoverability hypothesis implies that the set of explicit models constructed for counterfactuals will be greater in size than that constructed for comparable factual statements. For instance, consider the following two statements:

If Linda was in Dublin then Cathy was in Galway. (1)

If Linda had been in Dublin then Cathy would have been in Galway. (2)

Byrne and Tasso (1999) describe (1) as a conditional that deals with past facts and (2) as a conditional that deals with past possibilities. According to the authors,

Reasoners represent what is true in their models, and the conditional [(1)] is consistent with three separate situations that capture the way the world would be if it were true:

Linda	Cathy
not-Linda	not-Cathy
not-Linda	Cathy

—where the diagram uses Linda to represent Linda is in Dublin, Cathy to represent Cathy is in Galway, and not-Linda relies on a propositional-like tag for negation to represent that Linda is not in Dublin...” (p 727)

Byrne and Tasso (1999) further stated:

Reasoners construct an initial representation that is more economical than the fleshed-out set, because of the limitations of working memory:

Linda	Cathy
-------	-------

...

—where the three dots represent a model with no explicit content, which captures the idea that alternatives exist that have not been mentally articulated...Reasoners represent explicitly the case mentioned in the conditional, and they keep track of the possibility that there may be alternatives to it. (p. 727)

Thus, according to mental model theory, the abbreviated representation is determined, first, by the set of models that are judged to be possibly true and, second, by the framing of the conditional statement.

Although Byrne and Tasso stated that “reasoners engage in a similar process to understand the counterfactual conditional [(2)]” (1999, p. 728), they go on to propose that reasoners would represent (2) as follows:

factual:	not-Linda	not-Cathy
counterfactual:	Linda	Cathy

...

By the mental model account, the explicit counterfactual model is provided by the framing of the conditional statement much like the explicit factual model of (1) in the abbreviated representation. Unlike the case for representing (1), however, the factual model is somehow recovered, but the reason for this asymmetry is only weakly justified. For instance, Byrne and Tasso speculate that “the representation of the counterfactual may recruit memories that provide further information about the belief in the actual status of the antecedent, the consequent, and the connection between them” (1999, p. 728). The authors, however, do not explain why such recruitment processes cannot or do not operate in the “factual” case of (1).

I propose that a better theoretical account of the predicted difference in abbreviated representations of factual and counterfactual conditionals that take the form of (1) and (2) owes primarily to the level of epistemic uncertainty surrounding the statements and not to the factual-counterfactual distinction as mental model theory proposes. That is, conditionals that take the form of (1) do not strongly imply what actually happened, only what possibly could have happened if the conditional is true. A reasoner presented with (1) may be able to deduce the valid set of possible factual models but would not know which of them actually occurred. Conversely, it would also be impossible to know what possibilities actually did not occur.

In contrast, conditionals that take the form of (2) strongly imply that both the antecedent and consequent expressed in the statement certainly are false. It is as if (2) was an

abbreviated way of stating:

If Linda had been in Dublin then Cathy would have been in Galway. (3)

Linda was not in Galway.

Cathy was not in Dublin.

Thus, the factual model is recovered not because of a general rule of priority for representing factuais but because there is an implication of certainty that the antecedent and consequent referred to in the conditional in fact are false.

The “given that” framing used in the factual conditions of the present experiments may be compared with the “if” framing used in (1):

Given that Linda was in Dublin Cathy was in Galway (4)

Both (1) and (4) are factual conditionals, and both refer to the past. However, unlike (1), (4) greatly reduces the epistemic uncertainty about what actually occurred. It is as if (4) was an abbreviated way of stating:

If Linda was in Dublin Cathy was in Galway. (5)

Linda was in Dublin.

From this reading, only one possible factual model follows:

Linda	Cathy
-------	-------

Hence, a fairer test of whether factual and counterfactual conditional reasoning differ would involve comparing the factual “given that X was, Y was” and counterfactual “if X had been, Y would have been” constructions, both of which, I predict, will minimize epistemic uncertainty about what actually happened.

Compare (4) and (2), respectively:

Given that Linda was in Dublin Cathy was in Galway.

If Linda had been in Dublin then Cathy would have been in Galway.

Although the two conditionals are now well-matched in terms of the uncertainty that they convey concerning what actually happened, they still differ in terms of their gist. Whereas (4) implies the factual model

Linda	Cathy
-------	-------

(2) implies the factual model

not-Linda	not-Cathy
-----------	-----------

A better comparison of factual and counterfactual conditionals that is likely to maintain roughly equal level of uncertainty and to preserve gist might be the following two statements:

Given that Smith had high cholesterol, he had a heart attack. (6)

If Smith had not had high cholesterol, he would not have had a heart attack. (7)

Future research could profitably examine whether the conclusions drawn from factual and counterfactual conditionals such (6) and (7) differ reliably when combined with the same minor premise.

Directions for Future Research

The present research constituted the first set of studies that compared the effect of counterfactual and factual thinking focus on causal judgment. Although several precautions were taken and the results across these experiments were highly consistent, it is worth noting some of the ways in which future research may build on the present work. First, it will be important to measure causal judgment in a variety of ways. The present research relied on attitudinal assessments of causal statements like “I was the cause of the conflict.” Other approaches might involve having participants rate the impact of different causal candidates or having them list the causes that they deem relevant. Second, the range of dependent measures

can be extended to an investigation of other judgment, decision making, and reasoning processes, as well as to measures of emotion. If counterfactual thinking plays a critical role in amplifying emotions such as regret and disappointment, as Kahneman and Miller (1986) suggested, and if such thinking may attenuate other emotions such as anger and distrust, as Mandel (in press) recently reported, then we may expect to find a difference in emotional extremity as a function of whether people focus beforehand on counterfactual or factual assessments of a given episode. Research along these lines would undoubtedly help bridge the current gap between the psychology of “thinking about what might have been” and the psychology of “thinking about what was.”

References

- Branscombe, N. R., Owen, S., Garstka, T. A., & Coleman, J. (1996). Rape and accident counterfactuals: Who might have done otherwise and would it have changed the outcome? Journal of Applied Social Psychology, *26*, 1042-1067.
- Byrne, R. M. J. (1997). Cognitive processes in counterfactual thinking about what might have been. In D. L. Medin (Ed.), The psychology of learning and motivation, Vol. 37 (pp. 105-154). San Diego, CA: Academic Press.
- Byrne, R. M. J., & McEleney (2000). Counterfactual thinking about actions and failures to act. Journal of Experimental Psychology: Learning, Memory, and Cognition, *26*, 1318-1331.
- Byrne, R. M. J., & Tasso, A. (1999). Deductions from factual, possible, and counterfactual conditionals. Memory & Cognition, *27*, 726-740.
- 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.
- Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. Psychological Bulletin, *99*, 3-19.
- Gorovitz, S. (1965). Causal judgments and causal explanations. Journal of Philosophy, *62*, 695-711.
- Greenberg, J., Pyszczynski, T., & Solomon, S. (1982). The self-serving attributional bias: Beyond self-presentation. Journal of Experimental Social Psychology, *18*, 56-67.
- Hart, H. L. A., & Honoré, A. M. (1959). Causation in the law. Oxford, England: Clarendon.
- Hesslow, G. (1983). Explaining differences and weighting causes. Theoria, *49*, 87-111.
- Hilton, D. J. (1990). Conversational processes and causal explanation. Psychological Bulletin, *107*, 65-81.
- Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. Psychological Review, *93*, 75-88.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. Psychological Review, *93*, 136-153.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 201-208). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. American Psychologist, *39*, 341-350.
- Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. Journal of Personality and Social Psychology, *59*, 1101-1110.
- Lebow, R. N. (2000). What's so different about a counterfactual? World Politics, *52*, 550-585.

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. (in press). Counterfactuals, emotions, and context. Cognition and Emotion.

Mandel, D. R. (2001). Gain-loss framing and choice: Separating outcome formulations from descriptor formulations. Organizational Behavior and Human Decision Processes, *85*, 56-76.

Mandel, D. R. (2002). Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. Manuscript submitted for publication.

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.

McCloy, R., & Byrne, R. M. J. (2002). Semifactual "even if" thinking. Thinking and Reasoning, *8*, 41-67.

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.

Miller, D. T., & Ross, M. (1975). Self-serving biases in attribution of causality: Fact or fiction? Psychological Bulletin, *82*, 313-325.

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.

Miller, D. T., & Taylor, B. R. (1995). Counterfactual thought, regret, and superstition: How to avoid kicking yourself. In N. J. Roese & J. M. Olson (Eds.), What might have been: The social psychology of counterfactual thinking (pp. 305-331). Mahwah, NJ: Erlbaum.

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.

N'gbala, A., & Branscombe, N. R. (1997). When does action elicit more regret than inaction and is counterfactual mutation the mediator of this effect? Journal of Experimental Social Psychology, *33*, 324-343.

Roese, N. J. (1997). Counterfactual thinking. Psychological Bulletin, *121*, 133-148.

Roese, N. J., & Olson, J. M. (1993). The structure of counterfactual thought. Personality and Social Psychology Bulletin, *19*, 312-319.

Roese, N. J., & Olson, J. M. (1995a). Counterfactual thinking: A critical overview. In N. J. Roese & J. M. Olson (Eds.), What might have been: The social psychology of counterfactual thinking (pp. 1-56). Mahwah, NJ: Erlbaum.

Roese, N. J., & Olson, J. M. (1995b). Outcome controllability and counterfactual thinking. Personality and Social Psychology Bulletin, *21*, 620-628.

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.

Spellman, B. A. (1997). Crediting causality. Journal of Experimental Psychology: General, 126, 1-26.

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.

Tetlock, P. E. (1998). Close-call counterfactuals and belief system defenses: I was not almost wrong but I was almost right. Journal of Personality and Social Psychology, 75, 230-242.

Tetlock, P. E., & Belkin, A. (1996). Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives. In P. E. Tetlock & A. Belkin, A. (Eds.), Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives (pp. 3-38). Princeton, NJ: Princeton University Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-457.

Wasserman, E. A., Dorner, W. W., & Kao, S.-F. (1990). Contributions of specific cell information to judgments of interevent contingency. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 509-521.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. Journal of Personality and Social Psychology, 56, 161-169.

Author Notes

David R. Mandel, Department of Psychology.

This research was supported in part by a Research Development Grant from the University of Victoria. I thank Gail O'Halloran and Camelia Pakzad-Vaezi for their assistance with this research. I also wish to thank Mandeep Dhimi, Ahogni N'gbala, David Over, and two anonymous reviewers for their valuable feedback on drafts of this article.

Send correspondence 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.

Table 1

Correlations Among Judgment Ratings in Experiments 1 (bottom-left) and 2 (top-right)

Measure	Cause	Prevent	Blame	Control	$\underline{M}_{\text{EXP2}}$	$\underline{SD}_{\text{EXP2}}$
Causality	--	.43**	.50**	.35**	0.63	1.78
Preventability	.42**	--	.43**	.22*	0.89	1.77
Blame	.75**	.43**	--	.38**	0.28	1.73
Control	.33**	.22**	.39**	--	0.76	1.37
$\underline{M}_{\text{EXP1}}$	0.28	0.91	-0.13	0.28		
$\underline{SD}_{\text{EXP1}}$	2.02	3.05	2.04	2.01		

$\underline{df}_{\text{EXP1}} = 190$, $\underline{df}_{\text{EXP2}} = 94$, * two-tailed $p < .05$, ** two-tailed $p < .01$.

Table 2

Mean Rating on Causal Judgment Scale as a Function of Thinking Focus and Time in Experiment 1

Time	Thinking Focus		<u>M</u>	<u>SD</u>
	Counterfactual	Factual		
Recent	-0.52	0.66	0.11	1.64
Intermediate	0.94	0.41	0.73	1.76
Distant	0.43	0.01	0.19	1.69
<u>M</u>	0.31	0.36	0.33	1.71
<u>SD</u>	1.85	1.57		

Table 3

Mean Rating on Causal Judgment Scale as a Function of Thinking Focus and Person Focus in Experiment 2

Person Focus	Thinking Focus			<u>M</u>	<u>SD</u>
	Control	Counterfactual	Factual		
Self	0.66	0.50	0.30	0.48	1.20
Other	-0.14	1.36	1.16	0.79	1.25
<u>M</u>	0.26	0.93	0.72	0.64	1.23
<u>SD</u>	1.13	1.27	1.22		

Table 4

Mean Rating on Causal Judgment Scale as a Function of Thinking Focus and Action Focus
in Experiment 3

Action Focus	Thinking Focus		<u>M</u>	<u>SD</u>
	Counterfactual	Factual		
Omission	-0.72	-0.76	-0.74	1.20
Commission	-0.85	-1.24	-1.04	1.32
<u>M</u>	-0.78	-1.00	-0.89	1.27
<u>SD</u>	1.25	1.28		