Mental Simulation of Causality

Gary L. Wells
University of Alberta
Edmonton, Alberta, Canada

Igor Gavanski
Indiana University

We propose that people imagine alternatives to reality (counterfactuals) in assessing the causal role of a prior event. This process of mental simulation (Kahneman & Tversky, 1982) is used to derive novel predictions about the effects of default events on causal attribution. A default event is the alternative event that most readily comes to mind when a factual event is mentally mutated. The factual event is judged to be causal to the extent that its default undoes the outcome. When the boss was described as having considered another meal without the allergic ingredient, people were more likely to judge his decision as causal to the extent that the alternative event was said to have the allergic ingredient. In Experiment 2, a paraplegic couple was described as having died in an auto accident after having been denied a cab ride. People perceived the cabbie's refusal to take the couple as a stronger cause of the deaths than when the alternative meal was also said to have the allergic ingredient. We conclude that an adequate theory of causal judgment requires an understanding of these counterfactual simulations.

People evaluate an outcome in part by imagining alternatives to the outcome. For example, people might think that a man who broke his legs in an auto accident was lucky if the imagined alternative was death but think that he was unlucky if the imagined alternative was no accident at all. The mental construction of alternatives to factual events has been called counterfactual processing, and this type of processing has been implicated in certain judgments such as regret, perceived happiness, and victim compensation (Bell, 1982; Johnson, 1986; Kahneman & Miller, 1986; Kahneman & Tversky, 1982; Landman, 1987; Miller & McFarland, 1986). The purpose of this article is to provide evidence that counterfactual processing is basic to how people understand event causality.

The general proposition guiding this research is that an event will be judged as causal of an outcome to the extent that mutations to that event would undo the outcome. A mutation is a deletion, substitution, or other distortion of an event. The process of mutating events and assessing hypothetical alternative outcomes is part of the notion of mental simulation, a dynamic cognitive heuristic "that resembles the running of a simulation model" (Kahneman & Tversky, 1982, p. 201). The mental simulation heuristic is exemplified linguistically by if-only stems to verbal statements; for example, "if only the Watergate burglars had been more competent, then Gerald Ford would never have been President." The ability of people to mentally simulate counterfactual scenarios seems to be ubiquitous when people reflect on dramatic events, and we propose that counterfactual processing of this sort plays a central role in judging causality.

Mutability

For an event to be judged as causal, it must be psychologically mutable. For example, following a suicide in which a man leapt from a window, people would not cite the presence of gravity as a cause of his death. Although it is true that an absence of gravity would have undone the outcome, the presence of gravity is an immutable characteristic of life on Earth. Most events are not either mutable or immutable but instead vary in their degree of mutability. Recent research indicates that when people are asked to undo a dramatic outcome, exceptional events are more mutable than are normal events (Kahneman & Miller, 1986; Kahneman & Tversky, 1982; Wells, Taylor, & Turtle, 1987). Thus, when told that a tragic outcome occurred while a man was taking his usual route, people were less likely to mutate the decision about routes than when he was said to have taken an unusual route (Kahneman & Tversky, 1982; Wells et al., 1987). We would expect as well that people would cite the man's choice of route as more causal in the case where it was more mutable. The general proposition that mutability mediates per-
ceived causality seems to be an intuitively reasonable proposition.

Mutability is an important element in the mental simulation conceptualization of causal judgment processes. It might seem, therefore, that a test of the mental simulation conceptualization of perceived causality would involve manipulations to the mutability of events. In the experiments that we describe in this article, however, we did not manipulate mutability. This is because manipulations to the mutability of an event tend to be confounded conceptually with other possible explanations of the process. For example, we might manipulate mutability by manipulating the norm versus exception status of the event (Kahneman & Miller, 1986) or by manipulating whether the event was an action versus an inaction (Landman, 1987), both of which seem to affect mutability. However, other theories could explain the effects of these manipulations on causal judgments by invoking concepts such as consensus, distinctiveness, and salience. Thus, we do not believe that manipulations of mutability per se can distinguish between the mental simulation process and other, traditional views of the process.

A more convincing test of the mental simulation process requires manipulations to counterfactual events rather than manipulations to factual events. Our argument is that manipulations to counterfactual events should affect causal judgments only if people engage in counterfactual thinking. Previous research has not manipulated characteristics of counterfactual events even though this would seem to be the most direct way to see if people are engaging in a mental simulation of what might have been. In the next section we develop this idea of manipulating counterfactual events by introducing the concept of default events.

**Default Events**

A default event is a highly available counterfactual mutation to the factual event; it is the event that readily comes to mind as an alternative to the factual event that preceded some to-be-explained outcome. In some situations a default event is highly constrained. The alternative to the event "tails" in a coin flip, for example, is "heads," heads is a highly available default in part because it is exhaustive of all default values. The notion of default events is critical to understanding people's reactions to outcomes. Suppose, for example, you were choosing from Stock Options A and B. You choose Stock A, which later plummets, resulting in a $50,000 loss. Your reaction to this outcome is likely to be quite different if Stock B also plummets than if Stock B rises dramatically. In this case, choosing Stock B is the default event, and of course, the subsequent value of Stock B in no way affects your wealth (because you did not buy it). Nevertheless, it is assumed that the default event would undo the outcome if the value of Stock B rises and would not undo the outcome if the value of Stock B declines. In short, evaluation of the scenario depends not only on reality but also on what might have been in an assumed or imagined alternative to reality. In this stock option example, we would argue that the decision to buy Stock A would be perceived as more causal of the negative outcome (monetary loss) in the case where Stock B rises than in the case where Stock B plummets. Although it is true that the mental simulator could mutate toward some other event (such as a bond purchase instead of a stock purchase), the problem is framed to make the Stock A versus Stock B option a "joint" in the scenario (see Kahneman & Tversky, 1982, p. 207).

Now consider the following true story ("Cab’s Refusal Cited," 1986):

A young paraplegic couple requested a cab to take them to a nearby restaurant for dinner. When the cab arrived an hour later, the driver refused to take them. The couple decided to take their own car, which was equipped with special hand controls, to the restaurant. As they drove over a hill, the car slid off the road and struck a tree before landing in a ditch. The woman was killed and her husband spent a month in the hospital recovering from injuries incurred in the accident.

In this story there was one factual event (cab driver does not take the couple) that seems to stand out as mutable (probably because of its exception status). The default event is constrained by the presentation of the facts so that the alternative to the driver not taking the couple (driver taking the couple) is predicted to be the critical joint in the mental simulation of causality. If the default event (driver takes couple) undoes the outcome (the deaths), then the factual event will be judged as causal of the outcome. Indeed, some notion that the cabby caused the deaths seems to have been assumed by the headline writer of the story, which was headlined "Cab’s Refusal Cited in Death." Presumably, the cab driver is perceived to have had a role in the death because the reader mentally mutates the driver’s decision (from not taking the couple to taking the couple) and mentally simulates an alternative outcome (couple arrives safely). Although the mental simulator could conceivably mutate in a way that does not undo the outcome (e.g., driver takes couple and all die in a fiery crash), such a simulation would violate the general principle of normality. We return to the cab driver problem in Experiment 2.

It is important to note that scenarios can be manipulated so that the default event either does or does not undo the outcome. Suppose, for example, Mrs. Jones puts her daughter on Bus A, whereas she usually puts her daughter on Bus B. If Bus A has a wreck resulting in injury to Mrs. Jones’s daughter, then Bus B is a counterfactual default event and Mrs. Jones’s decision is seen as having played a role in her daughter’s misfortune. Suppose, however, Bus B also crashes. Now the counterfactual default event (if only Mrs. Jones had put her on Bus B) fails to undo the outcome.

If the counterfactual default event (putting the daughter on Bus B) does not change the outcome, then the factual event (putting her on Bus A) loses much of its causal significance. The attributor remains free to mutate mentally to other alternatives (such as not putting her on a bus at all), but the scenario outcome is more easily undone by the situation in which Bus B does not crash because mutations tend toward the direction of normality. In other words, the mental simulator is free to remove the offending outcome in both versions by imagining that the mother had not put the child on a bus at all; however, the version where Bus B does not crash is an easier mutation (because it is the norm), whereas the version where Bus B also crashes leads to an unsuccessful undoing of the outcome via mutation from Bus A to Bus B.

It is through the manipulation of the success-failure value of default events that our conception of the role of mental simulation in causal attribution departs from traditional theories of
causal attribution. There is nothing in traditional attribution theory that would predict the kinds of effects that we report in this article; our model of the mental simulation process, on the other hand, readily predicts these effects. The principal distinction between traditional attribution theories and our mental simulation model is that the former restrict the assumed processes to given facts or an assumed reality, whereas mental simulation involves a comparison between reality and what might have been. For example, traditional attribution theory might consider whether or not Mrs. Jones’s decision to put her daughter on Bus A was distinctive, consensual, or consistent (as in Kelley, 1972). However, traditional attribution theory does not require a consideration of what might have been and has no basis for making a prediction here. In other words, traditional conceptions of perceived causality are neutral or silent in predicting the effects hypothesized in these two experiments.

Although previous theorists have postulated a role for counterfactual processes in perceived causality (e.g., Einhorn & Hogarth, 1986; Kahneman & Miller, 1986; Mackie, 1974), our work develops beyond theirs in two important ways. First, we have moved beyond the mere proposition that mutability affects causal judgments to propose that the specific characteristics of the counterfactual event (default event) play an important role in the causality judgment. Second, previous treatments of the role of mental simulation in judged causality have not included empirical tests but instead have merely tried to make a conceptual link between mutability and causality.

Experiment 1

In Experiment 1, subjects read one of two versions of a story concerning a young woman, Karen, who went to a restaurant with her boss, Mr. Carlson. Mr. Carlson, unaware that Karen was allergic to wine, ordered a dish for her that contained a wine sauce. Karen ate the dish and had a severe allergic reaction from which she died. Mr. Carlson’s behavior and the outcome were the same in all conditions; that is, he ordered the wine dish and Karen died. The only difference is that in the one-wine story, Mr. Carlson first considered ordering an alternative dish for Karen that did not contain wine sauce, whereas in the two-wine story the alternative dish he considered also contained wine sauce. The story is structured so that the default event was the other meal considered by Mr. Carlson. The critical attribution process is hypothesized to occur when the mental simulator considers the alternative to reality (i.e., if only Mr. Carlson had ordered the other dish) and then assesses whether it undoes the outcome. Only the one-wine story satisfies the simulation heuristic’s criterion for assigning causality to Mr. Carlson’s behavior. If causal judgments are made at least in part by the mental simulation of alternative events, then subjects should assign more causality to Mr. Carlson’s ordering of the dish in the one-wine condition than in the two-wine condition.

The mental simulator could, of course, imagine alternatives other than the default event. For example, people could say, “if only Mr. Carlson had taken Karen to a Mexican restaurant” or “if only Karen had told him about her allergy.” However, these alternatives are as available in the one-wine condition as they are in the two-wine condition. The fact that other counterfactual simulations can and will come to mind does not negate the relative difference in the ease with which the outcome (Karen’s death) can be undone in the one-wine versus the two-wine condition.

We used two kinds of causal judgment measures in the first experiment. Some subjects made standard Likert-type judgments regarding the causal role of Mr. Carlson’s ordering decision, whereas others were asked to list the four most important causes of Karen’s death. We used both causality measures because of the possibility that manipulations to default events would affect the availability of the target event (as measured by whether the target event is mentioned among causes) but not causal judgments (when the subjects’ attention was focused specifically on the target event as happens with Likert-type, directed judgments). In addition, we were interested in subjects’ choices of mutations when they were asked to change the scenario so as to undo Karen’s death. We counterbalanced the order of the causal listing and mutation tasks. Finally, we asked subjects to explain why they assigned cause the way they did. This was an open-ended question asking them what they took into consideration when they made their causal judgments. We included this cause-explanation question only for qualitative discussion purposes and not for formal statistical analysis.

Method

Subjects. The sample was composed of 104 male and female introductory psychology students who participated in partial fulfillment of a course requirement.

Procedure. Subjects participated in groups of 5–20. They were randomly assigned to experimental conditions at the beginning of the session. Each subject read one version of the following story. The two-wine version contained the portions in parentheses; the one-wine version the portions in brackets.

Karen was an assistant editor for a small publishing firm. She had a rare hereditary disease called Karpinson’s hemotrysoma, characterized by the lack of an enzyme that normally breaks down certain proteins in the digestive system. Because of this, fermented drinks such as wine and liqueur can cause a severe allergic reaction in someone with the disease.

Karen had just received a promotion so her boss, Mr. Carlson, took her to an expensive French restaurant to celebrate. Mr. Carlson had been to this restaurant several times, so he ordered for both of them. As he looked over the menu, Mr. Carlson considered what to order for Karen. He first thought about ordering the Coquilles Saint-Jacques (Scallops Meuniere), but at the last moment decided on the Moules Marinieres instead. Although Mr. Carlson did not know this, both of these dishes were made with a wine sauce. (The Moules Marinieres was made in a wine sauce whereas the Scallops Meuniere did not contain any wine.)

Karen enjoyed her meal greatly, but began to feel ill shortly after finishing. Within minutes, she went into convulsions and was rushed away in an ambulance. She died on the way to the hospital.

The main dependent measures were of three kinds. Some participants were asked to list four things that could have been different in the story to prevent Karen’s death (mutation task) and to list the four most im-

---

2 Jones and Davis’s (1965) correspondent inference theory posits a role for effects of nonchosen alternatives, and in one sense, a nonchosen alternative is a default event or a version of what might have been. However, correspondent inference theory is a theory about the inference of traits and motives rather than event causality, and it is predicated on the assumption that the actor can foresee the consequences of chosen and nonchosen events.
Table 1

Causal Ratings, Percentage of Mutations, and Percentage of Causal Listing for the Ordering Decision in Experiment 1

<table>
<thead>
<tr>
<th>Default manipulation</th>
<th>Causal rating</th>
<th>Mutations</th>
<th>Causal listing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mutate first</td>
<td>List causes first</td>
<td>Mutate first</td>
</tr>
<tr>
<td>One wine</td>
<td>4.71</td>
<td>35.3%</td>
<td>38.9%</td>
</tr>
<tr>
<td>n</td>
<td>17</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>Two wine</td>
<td>3.29</td>
<td>0%</td>
<td>23.5%</td>
</tr>
<tr>
<td>n</td>
<td>17</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

Note. Percentages are of subjects who undid Karen’s death by mutating Mr. Carlson’s decision to order Meal A. Higher numbers on the causal rating indicate greater causality attributed to Mr. Carlson’s ordering decision.

Discussion

Subjects attributed greater causal significance to an event if its counterfactual default alternative would have yielded a different outcome than if the default alternative yielded the same outcome. This effect was obtained even though the event in question (Mr. Carlson’s choice of dishes) was identical in both conditions (ordering the Moules Mariniere), the consequence was identical in both conditions (Karen’s death), and the unforeseeability of the consequence was equal in both conditions (consequence unforeseen). The key difference seems to be contained in the if-only thinking inherent in mental simulation; participants in the one-wine condition could say “if only Mr. Carlson had ordered the other dish, then Karen would be alive today,” whereas participants in the two-wine condition could not.

The percentage of subjects who freely listed the target event as a cause in the one-wine condition was robustly higher than in the two-wine condition. Importantly, the default manipulation also affected the perceived causal magnitude of the target event (as measured on a Likert-type scale). Default manipulations, therefore, affect not only the availability or likelihood of the target event being mentioned among the important causes (as measured by having subjects list causes) but also subjects’ judgments of the target event’s causal significance.

Experiment 1’s results provide good evidence that mental simulation in general, and default event considerations in particular, are elementary to people’s understanding of causality. There is an alternative interpretation that we must consider, however. Specifically, it is possible that people thought that Mr. Carlson had less freedom of choice in the two-wine condition than he did in the one-wine condition. Being subject to external constraints (such as the degree of choice on the menu) might have reduced the perceived causal role of Mr. Carlson’s decision such that it is not construed as a decision at all, but as a mere reaction to the constraints. If so, then more traditional conceptions of attribution might accommodate these results.

There are two reasons why we do not favor the perceived-freedom interpretation of these results. First, subjects did not report anything akin to a perceived-freedom interpretation when asked why they made their causal attributions. For example, no subject said “having only two meals to choose from, he had no choice but to order a dish with wine sauce,” “his choice was restricted by the menu,” or anything remotely close to a choice-limitation or perceived-freedom interpretation. On the other hand, subjects commonly made statements such as “if only he had ordered the other dish,” “his choice was restricted by the menu,” or anything remotely close to a choice-limitation or perceived-freedom interpretation. On the other hand, subjects commonly made statements such as “if only he had ordered the other dish,” “his choice was restricted by the menu,” or anything remotely close to a choice-limitation or perceived-freedom interpretation.
perceived-freedom rationale in contrast to a prevalent use of if-only statements. A second reason that we prefer the mental simulation interpretation over the perceived-freedom interpretation is that the fact that mutations and causes were significantly related as indicated by the contingency analysis. Nevertheless, we cannot unequivocally rule out the perceived-freedom interpretation in the restaurant scenario. Thus, we chose to manipulate the effects of a default event in a second experiment for which we think such an alternative interpretation would be impossible or at best strained.

Experiment 2

In Experiment 2, participants read a story based on the true incident described earlier involving the couple and taxi driver. In both versions of the story, the taxi driver refused to take the couple so they decided to take their own car. They drove off a bridge, which had collapsed just shortly before, and both died. In one version of the story, the taxi driver reached the bridge shortly before it collapsed and made it safely across. In the other version, he also drove off the bridge and barely managed to escape his taxi before it sank beneath the water. Again, both the target event (the taxi driver’s refusal to take the couple) and the outcome (the couple’s deaths) are identical in both versions of the story. The difference is that the counterfactual default event (driver agrees to take couple) successfully undoes the outcome in one version but not in the other version. Unlike Experiment 1, there is no question about whether or not the two versions of the story differ in the perceived freedom of choice for the cabby; they do not.

In this experiment we did not have subjects list causes. Instead, all subjects made causal ratings and listed mutations. We varied the order of the causal rating and the mutation listing to test whether the default manipulation effect would be stronger if subjects mutated events prior to making these judgments than if the judgments preceded the mutation task. Such a pattern would be consistent with our contention that mutations mediate these judgments. We then asked subjects what they took into consideration in making their causal judgment measure. This is not because we consider causality and responsibility to be synonymous concepts (see Footnote 1), but because we suspected that most people would think that the driver’s behavior was at least slightly inappropriate. Thus, even though the driver did not intend or foresee the outcome, he is not blameless.

Method

Subjects. The sample was composed of 90 male and female introductory psychology students who participated in partial fulfillment of a course requirement. The study was a 2 (story version) × 2 (mutate first vs. judge cause first) between-subjects factorial design. There were between 19 and 26 participants per condition.

Procedure. Each participant read one of two versions of the following story. The cab-not-in-accident version contained the portion in parentheses, the cab-in-accident version the portion in brackets.

Eugene and Tina were a young married couple who lived in the country. Both were partially paralyzed and confined to wheelchairs. They had met four years before when Tina was a counselor with the Canadian Paraplegic Association, had fallen in love, and were married one year later.

On this particular evening, Eugene had phoned to request a cab to take them downtown. When the taxi driver arrived, Eugene and Tina were waiting by the street. On seeing that they were both in wheelchairs, the taxi driver refused their fare because he thought it would be too crowded in the taxi with both of them and the wheelchairs. So the taxi driver headed back downtown without them.

Because there was no time to call another cab, Eugene and Tina took Tina’s car, which was equipped with special hand controls. In order to get downtown from their house, they had to travel across a bridge over Rupert River. A severe storm the night before had weakened the structure of the bridge. About 5 minutes before Eugene and Tina reached it, a section of the bridge collapsed. (The taxi driver had reached the bridge about 15 minutes before them, and made it safely across.) [The taxi driver had reached the bridge shortly before them, and had driven off the collapsed bridge. He barely managed to escape from his taxi before it sank in the river.]

In the dark, Eugene and Tina drove off the collapsed bridge and their car plummeted into the river below. They both drowned. Their bodies were retrieved from the car the next morning.

After reading this story, all participants performed the mutation task, in which they listed four things that could have been different so that Eugene’s and Tina’s deaths might have been avoided. Participants also rated on a 9-point scale to what extent the taxi driver’s refusal to take Eugene and Tina caused their deaths (0 = not at all causal, 8 = very causal). Half the participants performed the mutation task first and half made causal ratings first. Participants then rated on a 9-point scale how responsible the taxi driver should feel for Eugene’s and Tina’s deaths (0 = not at all responsible, 8 = very responsible). They were then asked to report what they had taken into consideration in making their causal rating. Students were fully debriefed on completion of these tasks.

Results

We conducted separate analyses on the three measures, that is, ratings of the causal role of the driver’s decision, ratings of the driver’s responsibility in the deaths, and frequencies of mutations to the driver’s decision. Two judges coded the mutations independently and blind to condition according to whether a subject mutated the taxi driver’s decision to take the couple. Interjudge agreement was 100%. Table 2 presents the mean causal ratings, mean responsibility ratings, and percentages of mutations to the driver’s decision as a function of the default manipulation (cab not in accident vs. cab also in accident) and the order of the measures (mutate, then judge cause and responsibility vs. reverse order).

The means and percentages in Table 2 indicate that subjects rated the driver’s decision as more causal, rated the driver as more responsible, and mutated the driver’s decision more often in the conditions where the cab was not in an accident than in conditions where the cab was in an accident. An analysis of variance (ANOVA) on the causal ratings indicated that the main effect for the default manipulation was significant, $F(1, 86) = 11.01, p < .001$. We also conducted a contrast analysis of the form consistent with our hypothesis that the effect of the default manipulation on causal judgments would be stronger in the mutate-first condition than in the cause-first condition. This form of the predicted interaction was significant, $F(1, 86) = 16.363, p < .001$.

An ANOVA on the responsibility ratings produced a similar
pattern of significance. There was a main effect for the default manipulation, $F(1, 86) = 14.61, p < .001$, that was qualified by the predicted form of the interaction (stronger effect in the mutate-first than in the cause-first conditions) as tested by a contrast analysis, $F(1, 86) = 18.22, p < .001$.

Mutations were analyzed with a chi-square analysis of frequencies. Interestingly, the percentage of subjects who mutated the target event (driver's decision) was unaffected by whether they judged the causality before versus after making their mutations (55.4% in the cause-first condition and 57.2% in the mutation-first condition). Importantly, however, mutations to the target event were much more likely in the condition where the cab was not in an accident (84.4%) than in the condition where the cab was in an accident (28.2%), $\chi^2(1, n = 90) = 31.44, p < .001$.

Finally, we calculated correlations between the causal ratings, responsibility ratings, and ranks of mutations (where ranks were determined as in Experiment 1). Ratings of responsibility and causality were highly correlated, $r(88) = .702, p < .001$; cause ratings and mutations were also correlated, $r(88) = -.422, p < .01$, as were mutations and responsibility ratings, $r(88) = -.354, p < .01$. These latter two correlations were negative because we used ranks on the mutation measure (i.e., low mutation ranks indicate greater mutability).

**Discussion**

Subjects who were asked to describe event changes that could lead to a different outcome rated an event as more causal if its counterfactual default would have produced a different outcome than if its counterfactual default produced the same outcome. The effect was especially strong when subjects engaged in the undoing task prior to making their causal attributions. This interaction pattern provides strong support for our contention that mutations mediate causal judgments. It is important to note that the mutation task in no way directed subjects to make any particular mutation. Subjects could and commonly did mutate numerous events in the scenario (e.g., "if only the couple had taken a different route"). The difference in subjects' mutations to the target event (cab driver's decision) and their responsibility and causality judgments depended critically on the extent to which the default event was successful in undoing the deaths. The import of the default event for these judgments is therefore clearly demonstrated by these results.

One possible alternative interpretation of these results is that subjects in the cab also-in-accident condition somehow felt that the cab driver had "paid" for his causal role in the couple's deaths. Therefore, the subjects might have lowered their blame of the cab driver in terms of rating his causal role and responsibility lower in the cab also-in-accident condition than in the cab-not-in-accident condition. There are four reasons why we do not accept this alternative explanation (abbreviated as the cabby-punished interpretation). First, the .44 correlation between mutation ranks and causal judgments implicates the role of mutations in such judgments. We find it awkward at best to account for such a correlation using the cabby-punished interpretation.

A second reason for rejecting the cabby-punished interpretation is the Order X Default interaction. Why should the default effect on causal judgments be stronger when subjects mutate events prior to making the causal judgments? The cabby-punished interpretation seems to have no explanatory power for dealing with this interaction. A third reason for rejecting the cabby-punished interpretation is that such reasoning never surfaced in subjects' statements about why they judged cause the way they did. Common statements were "if the taxi driver had taken Eugene and Tina, they may not have died" or "it was the broken bridge, not the driver's refusal, that caused their deaths" or "they may have crashed in the taxi and died anyway" or "taxi driver could not have foreseen the tragic incident at the time of refusal." The most common statements were of the type we would expect if people were engaged in counterfactual simulation thoughts; none could be coded along lines consistent with the cabby-punished hypothesis.

The final reason why we do not favor the cabby-punished interpretation of Experiment 2 is that it has no general explanatory power across experiments. For example, this kind of interpretation does not explain the results of Experiment 1, whereas the mental simulation conceptualization not only explains both experiments but was used to derive clear a priori predictions. In short, the mental simulation framework explains more of the data than does the cabby-punished interpretation because mental simulation accounts for (a) the cause-mutation correlations, (b) the Order X Default interaction, (c) the causal statements made by subjects, and (d) the results of both experiments.

Thus, we argue that our subjects considered the cab driver's decision, mutated it toward a counterfactual default value (i.e.,
driver does take couple), and assessed whether the mutation could undo the outcome. The answer generally was no when the cab driver was also in an accident, but yes when the cab driver was not in an accident.

Judgments of responsibility followed a pattern that was almost identical to judgments of causality. Indeed, we see mental simulation as being critical to understanding how people make judgments of responsibility as well as judgments of causality. It should be noted, however, that we do not expect responsibility judgments always to follow the pattern of causality judgments as a function of default manipulations. Instead, we expect responsibility judgments to follow from default manipulations only if one or more additional conditions are met. These other conditions include, but are not perhaps restricted to, intention, foreseeability, or inappropriate actions. In Experiment 2, the cab driver’s action, refusing to take a paraplegic couple, was somewhat inappropriate even though there was no intention or foreseeability regarding the outcome. Within this background of an impropriety, the causal event in question seems to carry the additional weight of responsibility.3 Had we taken a measure of responsibility in Experiment 1, however, we would not have expected the default manipulations to have had much effect on responsibility judgments because none of the prerequisite conditions of intention, foreseeability, or inappropriate behavior characterized Mr. Carlson’s actions.

**General Discussion**

These two experiments provide persuasive evidence that causal reasoning is influenced by thoughts of what might have been. Although disclosure of a factual event might foster perceptions of its inevitability (as in the ubiquitous hindsight effect; Fischhoff, 1975, 1982), it can also evoke and direct a set of relevant counterfactual possibilities (Shafer, 1976). Learning of someone’s death surely evokes thoughts of how that death could have been avoided, and the set of counterfactual possibilities that comes to mind constitutes a sample of alternative events available for an assessment in the form of mental simulation. We introduced the concept of default events as a way of defining and controlling the set of counterfactual alternatives that people will enter into their mental simulation. The fact that manipulations to the counterfactual events influenced causal judgments indicates that counterfactual processing was involved in their judgment processes.

Although the effects predicted and observed in the two experiments described in this article might strike some readers as evidence of irrational or error-prone causal judgments, it is not our intention to make such an argument. We can construe the judgments predicted and observed in these experiments as being quite rational within the context of the processes that we postulate as mediators of those judgments. In general, we argue that the simulation heuristic does a good job of assigning causality to events in most cases. Because it is a heuristic it might be applied inappropriately in some circumstances, but we do not claim to have captured any evidence of erroneous judgments in these experiments.

Our introduction of the concept of a default event allowed us to separate the notion of mutability from the notion of undoing; some mutations can successfully undo an event and some cannot, depending on the characteristics of the default event. The concept of default events also highlights how our conceptualization differs from norm theory (Kahneman & Miller, 1986). In norm theory, mutable events are those that are nonnormal (exceptions), and mutations are always toward normality. In other words, norms are the defaults for exceptions. We disagree with this generalization for two reasons. First, we successfully manipulated default events in Experiment 1 without varying the normality of the events. Mr. Carlson’s choice of Moules Mariniere was no more or less normal than either Scallop’s Meuniere or Coquilles Saint-Jacques. Second, recent research indicates that mutations are not necessarily more likely to be made toward increased normality. Gavanski and Wells (in press) argued that the pervasive tendency in previous research for people to mutate toward normality owes to the fact that subjects were required to undo exceptional outcomes. Gavanski and Wells gave people outcomes that were either exceptional or normal and positive or negative and examined their mutation strategies when they were asked to undo the outcome. When the outcome was exceptional, people mutated events toward greater normality; when the outcome was normal, people mutated events toward greater exceptionality. Gavanski and Wells argued that people have a general belief heuristic; exceptional outcomes have exceptional causes and normal outcomes have normal causes. Thus, we propose that norms are not necessarily the default event of choice; if the person is trying to understand the cause of a normal outcome, then mutations tend toward exceptionality rather than normality. This is not meant to downplay the significance of norm theory as a general model of mental simulation. After all, we usually are interested in people’s judgments of exceptional outcomes (e.g., Why did John trip over Mary’s feet?) rather than normal outcomes (e.g., Why did John make it safely to his seat?) Furthermore, we suspect that exceptional outcomes are more likely than normal outcomes to evoke counterfactual processes naturally just as they are more likely to evoke spontaneous causal attributions (e.g., Yarkin, Town, & Harvey, 1981).

What would have happened if the outcomes in these experiments had been positive rather than negative? Gavanski and Wells (in press) found no effects for valence independent of the effects of normality on the nature of mutations that people made. Thus, if Karen had been the 1,000th person to have eaten Meal A and won a free meal and $50 for the honor, Mr. Carlson’s decision would have been perceived as causal of her new fortune according to the same mental simulation heuristics that led to judgments of causality for her death. Again, it would depend on the default event. If Meal B would not have won Karen the prize honors, then Mr. Carlson’s decision would have been judged as

---

3 There is an interesting parallel to this idea in many criminal and civil codes. Specifically, a person may be held criminally responsible for a death or injury that she or he neither intended nor foreseen (such as an accidental shooting) if she or he was engaged in an inappropriate act at the time (such as possession of a dangerous weapon). Similarly, a motorist who leaves his car unlocked with the key in the ignition is liable in some states for a victim who is run over by a thief who steals the car (Hart & Honore, 1985). Thus, formal elements of civil and criminal law acknowledge the idea that a relatively minor impropriety, without which the outcome would not have occurred, could lead to a judgment of responsibility for severe harm even if such harm was unforeseen and unintended.
causal; if Meal B would also have won her these honors, then Mr. Carlson’s decision would not have been judged as causal. On the other hand, recent research by Gleicher, Kost, Baker, Stratham, Richman, and Sherman (1988) suggests that spontaneous counterfactual processing may be more likely to occur in the context of negative than positive outcomes.

We found that our concept of default events was useful for generating some novel predictions about causal judgments. How does our description of mental simulation fit with traditional theories of attribution? We do not believe that the mental simulation conceptualization is incompatible with traditional theories. Indeed, some of the general concepts in traditional attribution theory, such as the general version of the covariation theories. Indeed, some of the general concepts in traditional attribution theory, such as the general version of the covariation theories. We argue, for example, that manipulations of distinctiveness information and consensus information (as in Kelley, 1972) are merely statements of what the attributor should mutate and how that mutation should affect the simulated outcome. For example, the statement “John failed the exam and others also did (did not) fail the exam” directs the attributor to mutate the identity of the actor and informs whether the mutation undoes the outcome (i.e., the failure on the exam). The mutation of John’s identity to a default identity (someone else in the class) successfully undoes the outcome in the case of low consensus but does not undo the outcome in the case of high consensus. Thus, John’s identity seems causal when its mutation undoes the outcome (low consensus) and not otherwise. A similar argument applies to distinctiveness information, except that distinctiveness information calls for a simulation of what happens when one mutates the situation (rather than the person). Consider the statement “John failed the exam and he fails (does not fail) exams in other courses.” This information directs the attributor to mutate the situation (this course) to another situation (another course) and informs whether the mutation undoes the outcome (failure). High-distinctiveness information calls for a particular solution to the mental simulation, namely, the mutation of situation does indeed undo the outcome, and therefore the situation is causal. The discounting and augmentation principles of causal attribution can be similarly understood as products of mental simulation whereby the person mutates (via deletion and addition of plausible causes) and assesses the extent to which the mutation undoes the outcome.

As with most research on causal reasoning, our subjects were given certain bits of information and we examined how and whether the information affected their causal judgments. The question arises as to whether or not people naturally use such information in a nonexperimental setting. Unlike previous studies on causal reasoning, however, our subjects used information that we did not provide, namely, imagined counterfactual or default information. In Experiment 2, for example, we did not provide a counterfactual nor did we tell subjects what effect the counterfactual would have had on the outcome. Interestingly, some readers of a previous version of this article have suggested that the scenario provided subjects with the counterfactual and its outcome. In fact, however, subjects had to generate the counterfactual (had the driver taken the couple . . . ) and simulate its outcome (the couple would [would not] have died). The mistaken impression that we provided subjects with the counterfactual and its outcome perhaps attests to the extent to which counterfactual possibilities and their simulated outcomes naturally come to mind. In addition, it should be noted that we obtained the default effect on causal judgments in both experiments even for subjects who were not instructed to engage in counterfactual thinking prior to making their causal judgments.

Another reason that we believe people naturally use counterfactual processing to evaluate outcomes stems from some real-life comments made by accident victims. Bulman and Wortman (1977) found that all of the severe accident victims in their research reported asking themselves “why me?” and found that people commonly gave counterfactual responses to this question. For example, their subjects said “if only I hadn’t ridden with her that day I wouldn’t be paralyzed” (p. 360) or “I always figure I could have taken the bus home, like I usually did” (p. 361). Indeed, we believe that the scenarios we used in our experiments are “natural stories” in that they are merely a sequence of events followed by a dramatic outcome, that such stories are common, and that such scenarios capture the way people typically organize event sequences. In fact, as noted earlier, the cab driver scenario was drawn from a true story as reported in a Canadian press article.

We propose that the mental simulation of counterfactuals is a “natural” analogue to thinking processes that underlie a diverse set of judgments that people make about social events. Judgments of regret or joy, for example, can be construed as resulting from the ease with which an alternative to a negative or positive outcome can be imagined by mentally mutating prior events toward their default counterparts. By natural analogue we mean that people seem actually to speak and think in ways that parallel counterfactual mental simulation. For example, people often explain failure by making an if-only statement (e.g., “if only I had taken Professor Jones’s class”) and people try to persuade others about judgments of fault by reference to if-only statements (e.g., “if only the police had not arrived in riot gear . . .”). Indeed, courts of law have tried to clarify for jurors the notion of cause by developing a but-for or sine qua non rule, to wit: “The defendant’s conduct is not a cause of the event, if the event would have occurred without it” (Prosser, 1971, p. 239). The tendency for people to use language in a way

---

4 In a recent attack on the notion of mental simulation, Rips (1986) argued that complete or error-free mental simulation of complex sequences “are off limits, given the normal memory and processing limits that humans have to contend with” (p. 271) and that mental modelers must therefore “forfeit the claim that ‘simulation’ should be taken literally” (p. 274). We agree with this assertion only insofar as mental simulation is assumed to involve a complex system of all possible states. We believe, as do Kahneman and Tversky (1982), that the envisioning process is selective and that the goal of mental simulation research and theory is to discover and define the psychological processes of selection. Only by knowing these selection processes can artificial intelligence programs based on mental simulation (e.g., Forbus’s, 1983, computer program on the behavior of balls bouncing in a two-dimensional plane) manage to stay within reasonable processing limits.
that parallels the mental simulation conceptualization ("if only . . .") and for tort law to spell out a similar judgment process suggests to us that the mental simulation idea is part of people's natural thinking processes. Anecdotally, we add the observation that introductory lectures about attribution seem to ring hollow for students when we discuss traditional attribution theories, whereas the mental simulation conceptualization seems to ring true to students as being "the thoughts that I go through when trying to decide what caused something." We do not suggest that self-reports of this type constitute convincing evidence that people's causal judgments are governed by the mental simulation heuristic, but we have been impressed with the extent to which mental simulation of counterfactuals seems to be a natural analogue to people's conscious thoughts.

We have shown that the mental simulation framework can predict effects on causal judgment for which traditional theories are silent. We find it difficult to account for people's judgments about Mr. Carlson's causal role in Karen's death in Experiment 1 or the cab driver's causal role in the couple's car accident in Experiment 2 without assuming that people were imagining an alternative to reality. These results, in conjunction with what we consider to be an eminently reasonable and intuitively appealing conception of how people think about causality, lead us to argue that an adequate theory of causal judgment requires concessions to and explorations of people's counterfactual processing of information.

References