Culpable Control and Counterfactual Reasoning in the Psychology of Blame

Mark D. Alicke
Ohio University

Justin Buckingham
Towson University

Ethan Zell
Ohio University

Teresa Davis
Middle Tennessee State University

Many counterfactual reasoning studies assess how people ascribe blame for harmful actions. By itself, the knowledge that a harmful outcome could easily have been avoided does not predict blame. In three studies, the authors showed that an outcome's mutability influences blame and related judgments when it is coupled with a basis for negative evaluations. Study 1 showed that mutability influenced blame and compensation judgments when a physician was negligent but not when the physician took reasonable precautions to prevent harm. Study 2 showed that this finding was attenuated when the victim contributed to his own demise. In Study 3, whether an actor just missed arriving on time to see his dying mother or had no chance to see her influenced his blameworthiness when his reason for being late provided a basis for negative evaluations but made no difference when there was a positive reason for the delay. These findings clarify the conditions under which an outcome's mutability is likely to influence blame and related attributions.

Keywords: *blame; counterfactual; causation; culpability*

Harmful events are fodder for attributional investigation. One reason why harmful events provoke attributional activity is that they are often unexpected, and people are prone to seek explanations for surprising events (Hastie, 1984; Pyszczynski & Greenberg, 1981; Wong & Weiner, 1981). A related reason is that harmful outcomes are unfavorable, and unfavorable outcomes elicit more causal reasoning than favorable ones (Boninger, Gleicher, & Strathman, 1994; Johnson, 1986). Yet another reason for scrutinizing harmful events is to avoid being victimized by them (Burger, 1981; Shaver, 1970). Finally, emotional reactions to a victim's suffering may impel observers to analyze the causes of the victim's misfortune.

Ordinary explanations of harmful events differ markedly from formal scientific explanations. Whereas scientific explanations identify universal laws that associate antecedent and consequent events, ordinary explanations devolve on specific episodes (Antaki & Fielding, 1981; Hart & Honore, 1959; Hilton & Slugoski, 1986). For example, a woman who was trying to understand why she was assaulted would be relatively uninformed by sociological explanations about economic decline. Rather, she wants to know why she was assaulted at a particular place and time. Causal candidates might include changing her route to work because she was in a hurry. Although route changes are not universally linked to assault, they provide the key to understanding this particular event. As Hart and Honore (1959) emphasize in their discussion of causation in the law, people explain harmful events by identifying abnormal conditions that

Authors' Note: Please address correspondence to Mark D. Allicke, Department of Psychology, Ohio University, Athens, OH 45701; e-mail: alicke@ohio.edu.

PSPB, Vol. 34 No. 10, October 2008 1371-1381
DOI: 10.1177/0146167208321594

differentiate the harmful event from innocuous ones (see also Hilton, 1990, 1991).

Counterfactual Reasoning and Harmful Events

Contrasting abnormal and normal conditions is a form of counterfactual reasoning (Kahneman & Miller, 1986; Mandel, Hilton, & Catellani, 2005; Roese, 1997; Roese & Olson, 1995). In the context of harmful events, counterfactual reasoning entails thinking about how deleterious outcomes could have been avoided. Roese (1997) has proposed two distinct stages of counterfactual generation. In the first stage, counterfactual reasoning is activated by affectively-arousing information, such as learning about a victim's being harmed. The second stage involves considering specific alternatives by which the harmful outcome could have been averted. Research suggests that in considering alternatives to unfortunate outcomes, observers focus on unusual, unexpected, or exceptional circumstances (Kahneman & Miller, 1986; Kahneman & Tversky, 1982; Miller, Turnbull, & McFarland, 1989; Roese & Olson, 1995).

An early study by Miller and McFarland (1986, Study 1) exemplifies counterfactual reasoning explanations of harmful events. Participants in this experiment learned that a man was shot and injured while visiting his usual convenience store or a different one. The primary response measure was the amount of money participants awarded to the victim. Results showed that more compensation was recommended when the man was shot in the unusual than in the usual store. The counterfactual reasoning explanation of these results can be understood in terms of Roese's two-stage process. First, harmful events evoke negative affect and instigate counterfactual investigations. In Miller and McFarland's scenario, the innocent victim's ordeal activates the counterfactual reasoning process. Second, the content of counterfactual reasoning is determined by the event's most mutable features. The distinguishing mutable feature in these scenarios is the victim's decision to visit a different convenience store. In this condition, counterfactual reasoning is expressed in ruminations such as "If only the actor had gone to his usual store, he wouldn't have been shot." Counterfactual thoughts presumably lead observers to sympathize with the victim and to award him more compensation.

Counterfactual reasoning theories, however, do not unambiguously predict Miller and McFarland's (1986) results. Counterfactual theories stipulate that observers will respond emotionally to unfortunate events and that they will seek to explain such events with reference to their mutable features. But counterfactual theories do not explain fully the link between mutability and blame. In particular, the assumption that the unusual store scenario will lead observers to sympathize more with the actor's plight is not a unique derivation. One could as easily assume that observers would blame the actor more for his willful decision to go to the unusual store and therefore award him less compensation. In this case, observers might reason counterfactually: "If only the victim had not made the stupid decision to go to the unusual store, he wouldn't have been harmed." (For a similar example, see Roese, 1997, p. 142.)

In fact, recent research by Goldinger, Kleider, Azuma, and Beike (2003) supports this possibility. They created scenarios in which a victim was injured at a basketball game after changing his seat or remaining in his original seat. In contrast to Miller and McFarland's (1986) findings, the victim's decision to move his seat, which increases his control over the outcome, led to *increased* blame and *decreased* compensation. Research by Mandel and Lehman (1996) suggests that other factors, especially perceived controllability and preventability, either supplant or work in conjunction with counterfactual cues such as normality or mutability to alter blame and causal judgments.

Culpable Control

What is needed, therefore, is a general basis for predicting when mutable event features will increase or decrease blame attributions. In this article, we enlist the "culpable control" model (Alicke, 2000) to bridge the gap from mutability to blame. According to the culpable control perspective, when observers review events, they consciously attend to the blame criteria that Anglo-American legal and moral traditions prescribe. This includes assessing the actor's causal influence over the harmful outcome, determining whether the outcomes were brought about intentionally and with foresight, and considering whether mitigating or extenuating circumstances existed (Shaver, 1985). These judgments pertain to how much control the actor exerted over the harmful outcomes.[1]

At the same time, observers spontaneously evaluate the desirability of each information item (Alicke, 2000). Spontaneous evaluations are favorable or unfavorable evaluative responses to the participants in the event, their actions, and the outcomes that eventuate. Observers, for example, may have a positive response to the perpetrator because she is socially attractive or because they approve of the values her behavior reflects, or a negative response because she belongs to a despised ethnic group or because her behavior causes disastrous outcomes. Ultimately, blame attributions are determined by the perception that the perpetrator controlled a harmful outcome in a way that arouses negative evaluative reactions.

The connection between culpable control and counterfactual reasoning in relation to harmful events is as follows: Counterfactual reasoning theorists stipulate that the knowledge that an outcome could have been averted amplifies observers' emotional responses to the event. According to the culpable control model, whether this heightened affect influences blame depends on whether an actor has exerted control in a way that arouses negative spontaneous reactions. If an actor has taken all reasonable precautions to avoid harm, and harm nevertheless occurs, the outcome's avoidability will have little effect on blame. In this case, there is no basis in the control evidence for negative evaluations, and so the heightened affect that the outcome's avoidability has aroused has nothing to "work on." But, when harmful outcomes result from negligence, recklessness, selfishness, or malevolence, mutability becomes an aggravating factor. In this case, the heightened emotional reactions that occur in relation to the knowledge that the harmful outcome was avoidable, coupled with a basis for negative evaluative reactions, amplifies blame beyond the level that would exist if the outcome had been unavoidable.

We conducted three studies to show that the effects of mutability on blame depend on whether there is a basis for negative evaluations and where the locus of those evaluations resides. Study 1 shows that a perpetrator is blamed more when the harmful effects of his actions are mutable, but only if he acted negligently. Study 2 replicates these findings when the victim is blameless, but shows that they are negated when the victim is a dislikable character who contributes to his plight. Study 3 shows that extra-evidential information that establishes a positive or negative motive and makes the perpetrator likable or dislikable determines whether mutability information influences blame judgments.

STUDY 1

In the first study, we depicted a situation in which a physician prescribed a drug for a woman who suffered from a life-threatening infection. The woman had a severe allergic reaction to this medication and died. In the mutable outcome condition, the doctor could have prescribed another drug that would have saved her life, whereas in the immutable outcome condition, the other medication would also have been lethal.

The central prediction in Study 1 involves the interaction of outcome mutability with information about the physician's negligence. When the physician takes all reasonable precautions to avoid harm, and there is no basis in the control evidence for negative evaluations of him or his behavior, the outcome's mutability should have no effect on blame. When the physician is negligent, however, and his behavior provides a basis for censure, the knowledge that the harmful outcome could have been avoided should lead to increased blame. This reasoning translates into an interaction between outcome mutability and negligence such that outcome mutability effects on blame and causation should emerge when the physician behaves negligently, but not when he takes all reasonable precautions to avoid harm.

Method

Participants, scenario, and design. Participants in this and subsequent studies were recruited from undergraduate psychology classes. We did not obtain any effects due to gender in these studies. Data in each study (except for Study 3) were collected in group-testing sessions ranging from approximately 25 to 75 participants per session.

One hundred two students participated in Study 1. The design varied whether a physician behaved negligently or nonnegligently and whether the harmful consequences were mutable or immutable. The complete scenario and experimental conditions are described below.

Common information. Dr. Richard Holmes was a physician employed in a large hospital in the Southwest where he was a specialist in infectious diseases. On May 27, 1997, Holmes was presented with the case of a woman named Karen Trollier. Mrs. Trollier was suffering from a life-threatening lung infection. In cases such as Mrs. Trollier's, this infection, which without treatment inevitably leads to death, can be cured completely by a drug called Propanil-B. Unfortunately, routine blood tests conducted before Mrs. Trollier was to receive treatment showed that she was allergic to Propanil-B. The lab informed Mrs. Trollier that the results of her tests would be sent to Dr. Holmes the next day.

Physician nonnegligent condition. Unfortunately, the lab sent the wrong results to Dr. Holmes. Thus, the report Dr. Holmes read stated that Mrs. Trollier had no known allergies. Nevertheless, to be on the safe side, Dr. Holmes checked again with the lab and was again given the wrong results. Apparently, the lab had mixed up Mrs. Trollier with another patient. Dr. Holmes, therefore, had no way of knowing about her allergy and prescribed the routine dosage of Propanil-B. Four hours after taking the medication, Mrs. Trollier had a severe reaction and died.

Physician negligent condition. Unfortunately, Dr. Holmes was very busy and did not read the lab results carefully. Thus, Dr. Holmes failed to recognize that Mrs. Trollier was allergic to Propanil-B. Dr. Holmes, therefore, prescribed the routine dosage of Propanil-B.

Four hours after taking the medication, Mrs. Trollier had a severe reaction and died.

Immutable condition. If Dr. Holmes had read Mrs. Trollier's chart correctly (negligent condition)/If the lab had given Dr. Holmes the correct information about Mrs. Trollier's allergies (nonnegligent condition), he could have prescribed a different drug called Ethericin. Ethericin has a number of undesirable side effects and is, therefore, usually recommended only for people who are allergic to Propanil-B. However, an autopsy conducted on Mrs. Trollier discovered that she had a rare liver disorder that would have caused her to react even more severely to Ethericin. Thus, Mrs. Trollier would have died regardless of which drug Dr. Holmes administered.

Mutable condition. If Dr. Holmes had read Mrs. Trollier's chart correctly (negligent condition)/If the lab had given Dr. Holmes the correct information about Mrs. Trollier's allergies (nonnegligent condition), he could have prescribed a different drug called Ethericin. Ethericin has a number of undesirable side effects and is, therefore, usually recommended only for people who are allergic to Propanil-B. However, Ethericin would almost certainly have saved Mrs. Trollier's life.

Dependent Measures

Blame ratings ("To what extent do you blame Dr. Holmes for Mrs. Trollier's death?") were made on 11-point scales ranging from 0 (*no blame at all*) to 10 (*a great deal of blame*). Victim compensation recommendations ("How much money, if any, do you think Trollier's family should be awarded from Holmes' insurance company?") were made on a 21-point scale ranging from $0 to $1,000,000 in increments of $50,000. Negligence ratings ("To what extent do you think Dr. Holmes behaved negligently?") were included as a manipulation check on the negligence manipulation (0 = *not at all negligent;* 10 = *extremely negligent*). Finally, ratings of the perpetrator's causal control ("To what extent do you think that Dr. Holmes' actions were a cause of Mrs. Trollier's death?") were made on 11-point scales (0 = *not at all a cause of the patient's death;* 10 = *very strong cause of the patient's death*).

Results

Table 1 displays means and standard deviations for ratings of negligence, blameworthiness, causation, and compensation recommendations.

The manipulation check on negligence showed, as intended, that the physician was judged to be more negligent when he behaved negligently than when the lab was negligent, $F(1, 96) = 361.93$, $p < .0001$. Significant main effects were also obtained for the negligence variable on ratings of the physician's perceived blameworthiness, $F(1, 96) = 205.21$, $p < .0001$, and causal control, $F(1, 96) = 72.51$, $p < .0001$, and on compensation recommendations, $F(1, 96) = 23.15$, $p < .0001$. As would be expected, the physician was seen as more blameworthy and more causal, and the victim's family as deserving of more compensation, when he behaved negligently versus when the lab was negligent.

TABLE 1: Means and Standard Deviations for Blameworthiness, Causation, Negligence, and Compensation (Study 1)

	Physician Negligent	*Physician Nonnegligent*
Negligence		
Mutable outcome		
M	8.92	1.68
SD	1.44	2.81
Immutable outcome		
M	8.81	1.72
SD	1.44	1.59
Blameworthiness		
Mutable outcome		
M	9.00	1.56
SD	1.23	2.40
Immutable outcome		
M	6.08	1.36
SD	2.70	1.95
Causation		
Mutable outcome		
M	7.58	1.48
SD	2.60	2.39
Immutable outcome		
M	4.38	1.40
SD	3.10	1.65
Compensation		
Mutable Outcome		
M	16.38	7.40
SD	6.06	6.79
Immutable outcome		
M	8.31	5.72
SD	5.63	5.76

We anticipated parallel effects of outcome mutability. This expectation was confirmed on ratings of blame, $F(1, 96) = 13.54$, $p < .0004$, and causation, $F(1, 96) = 12.97$, $p < .0005$, and on compensation recommendations, $F(1, 96) = 16.46$, $p < .0001$. As Table 1 shows, the physician was blamed more, seen to have had more causal control, and required to pay more compensation to the patient's family when the patient's death was avoidable versus unavoidable. Outcome mutability did not influence negligence ratings ($F < 1$).

The main focus of Study 1 was on the predicted interaction between outcome mutability and negligence: We expected that outcome mutability effects would be stronger when the physician behaved negligently than when he took all reasonable precautions to prevent

harm. This interaction was obtained on ratings of blame, $F(1, 96) = 10.30$, $p < .002$, and causation, $F(1, 96) = 7.08$, $p < .01$, and on compensation recommendations, $F(1, 96) = 7.07$, $p < .009$. Simple effects tests revealed the predicted pattern: Outcome mutability exerted a significant influence on blame, $F(1, 96) = 22.77$, $p < .0001$, and causation, $F(1, 96) = 19.31$, $p < .0001$, and on compensation, $F(1, 96) = 21.23$, $p < .0001$, when the physician was negligent but not when the physician was nonnegligent (each $F < 1$).

An alternative explanation to the culpable control account of these findings is to assume that the negligence manipulation concomitantly increases the perpetrator's causal effect on the outcome. According to this view, increased perceptions of causation, rather than negative evaluations of the physician's behavior, can account for our findings. A purely causal account assumes that mutability provides one source of causal information in that a perpetrator has a larger causal influence when an outcome is avoidable versus unavoidable, and negligence provides a second source of causal influence in that the negligent physician is more causal than the nonnegligent one. When the outcome is immutable, the actor cannot change the outcome, and so only his negligence influences blame. When the outcome is mutable, both negligence and mutability combine to influence blame.

One problem with this additive causal model is that it fails to explain why the mutability effect is negated in the nonnegligent physician condition. If negligence and mutability were simply two additive sources of causal influence, then when the physician was nonnegligent, mutability would independently influence blame. The fact that there was no effect of mutability when the physician was nonnegligent works against the purely causal view. Nevertheless, we conducted two separate covariance analyses to assess this interpretation. The first analysis statistically removed causal ratings from the negligence main effect, and the second analysis removed causal ratings from the Negligence × Mutability interaction. The results of this analysis showed that the negligence main effect remained significant after covarying out participants' ratings of the perpetrator's causal influence on the outcome, $F(1, 95) = 75.50$, $p < .0001$, as did the interaction effect, $F(1, 95) = 3.88$, $p < .05$. Thus, although causal judgments are undoubtedly important here, they do not completely explain the results of Study 1. The findings are consistent with our assumption that an outcome's mutability influences blame attributions only when coupled with a basis for negative evaluations, which in this instance were represented by a physician's negligent behavior.

STUDY 2

We applied culpable control reasoning in Study 2 to incorporate the role of both perpetrators and victims in harmful events. In particular, we tested the assumption that the pattern of perpetrator attributions we obtained in Study 1 would be moderated by an unattractive victim's contribution to his own demise earlier in the sequence of events. We used a different scenario in Study 2 to extend the generality of the findings. In this scenario, a patient who had been shot died during a surgical procedure for which the surgeon was either negligent or nonnegligent. In some conditions, the patient would have died regardless of the surgeon's negligence, whereas in the other conditions, his life would have been saved. The new variable, victim characterization, was manipulated by presenting some participants with a positive characterization in which the victim was described as a businessman who was shot while trying to resolve a quarrel, and others with a negative characterization in which the victim was a former prisoner who was shot while threatening a policeman.

Our reasoning about the interplay among perpetrator negligence, outcome mutability, and victim characterization was as follows: When the physician takes all reasonable precautions to prevent harm, his blameworthiness should be affected minimally by the outcome's mutability or the victim's characterization. When the physician is negligent, as in the previous study, he should be blamed more for a mutable than for an immutable outcome. This effect should be qualified, however, by whether the victim's objectionable behavior contributes to his own misfortune. We believe that negative reactions to a physician who causes an avoidable death will be attenuated when the victim contributes to the calamity. Thus, the difference between attributions to the negligent doctor in the mutable and immutable conditions should be larger when the victim's actions are characterized positively rather than negatively.

Method

Participants, scenario, and design. Participants in Study 2 were 270 undergraduate students. As in the previous study, the design varied whether a physician behaved negligently or nonnegligently and whether the harmful consequences were mutable or immutable. In addition, the victim characterization was positive or negative. The complete scenario and experimental conditions are described below.

Common information. On March 7, 1994, Steven J. Langsford died on the operating table during a surgical

procedure performed by Dr. Robert Fallsworth. The following facts are undisputed:

1. At the time of the surgery, Langsford was 37 years old and had never had any major health problems.
2. Langsford lived in a suburb of Chicago, Illinois.

Positive victim characterization

3. Langsford was the foreman of a construction team.
4. Langsford was in Chicago to meet with some people who wanted to discuss a building project.
5. Langsford's meeting lasted for about an hour, after which he went to the hotel to meet with his wife.
6. Langsford's wife was shopping with friends, so he watched TV for about an hour.
7. At about 6:30 p.m., Langsford and his wife went to have dinner in the hotel.
8. Langsford and his wife saw some acquaintances and went over to their table to say hello.
9. As they were walking toward the table, Langsford saw a couple in the corner engaged in a violent argument.
10. When Langsford saw the man grab the woman's arm roughly, he walked over to see if he could help.
11. The man pulled a gun and shot Langsford twice.

Negative victim characterization

3. Langsford had recently been released from a state prison where he had served a 7-year sentence for armed robbery.
4. Langsford was in Chicago to meet with three former inmates of the prison.
5. Langsford's meeting lasted for about an hour. One of the former inmates later reported that Langsford had proposed a drug deal but that nobody was interested.
6. Langsford saw a woman he used to date and asked her if he could buy her a drink.
7. The woman, who disliked Langsford, politely said no and said she was waiting for a friend.
8. Langsford bought her a drink anyway, and when the woman got up to move her seat, Langsford roughly grabbed her arm and kept her from leaving.
9. The manager of the bar called the police, who happened to be patrolling close to the area of the bar.
10. Langsford saw the policeman before the policeman saw Langsford, and Langsford immediately reached for a gun that he had concealed in his coat pocket.
11. Luckily, the policeman's partner, who was still at the doorway, saw Langsford reach for the gun and shot Langsford just as he was about to fire at the other policeman.

Common information

12. An ambulance brought Langsford to the nearest hospital.
13. Dr. Fallsworth, the thoracic surgeon on call, performed an emergency operation.
14. Langsford died on the operating table as a result of having lost a great deal of blood.
15. During the hospital's routine investigation following a patient's death, a physician who read Langsford's chart learned that Langsford had a blood disorder that required a special medication before surgery to prevent excessive bleeding.
16. Two other doctors who looked over the chart confirmed the physician's report.

Physician nonnegligent

17. An investigation by the hospital showed that Fallsworth had read Langsford's chart and had ordered the proper blood coagulant before performing the surgery to prevent excessive bleeding. Fallsworth had even called the lab twice to remind them to send the correct blood coagulant.
18. The investigation showed, however, that the blood lab made a serious mistake by sending a blood thinner and mislabeling the package.

Physician negligent

17. An investigation by the hospital showed that Fallsworth had failed to read the part of the patient's chart that mentioned the blood disorder.
18. The investigation also showed that a resident had reminded Fallsworth that he had not finished reading Langsford's chart but that Fallsworth ignored him.

Mutable outcome

19. The investigation further showed that if the correct blood coagulant had been administered, Langsford would definitely have survived the operation.

Immutable outcome

19. The investigation showed, however, that even if the correct blood coagulant had been administered, Langsford had already lost too much blood before he got to the hospital and could not possibly have survived the operation.

Response measures. The same blameworthiness, causal, and negligence questions were asked and rated on the same scales as in the first study, the only changes being in the names of the perpetrator and victim. In this study, victim compensation was measured on an 11-point scale ranging from 0 = *far less than the average amount in such cases* to 10 = *far more than the average amount in such cases*.

Results

Table 2 displays means and standard deviations for ratings of negligence, blameworthiness, causation, and compensation recommendations.

The manipulation check on the physician's negligence showed, as intended, that the negligent physician

TABLE 2: Means and Standard Deviations for Blameworthiness, Causation, Negligence, Punishment, and Compensation (Study 2)

	Attractive Victim Characterization		*Unattractive Victim Characterization*	
	Physician Negligent	*Physician Nonnegligent*	*Physician Negligent*	*Physician Nonnegligent*
Negligence				
Mutable outcome				
M	7.94	2.56	6.79	2.94
SD	2.59	2.67	3.10	3.33
Immutable outcome				
M	7.50	2.06	6.97	1.31
SD	2.62	2.90	3.19	2.35
Blameworthiness				
Mutable outcome				
M	7.09	2.06	5.00	3.09
SD	2.80	2.50	3.34	3.31
Immutable outcome				
M	4.06	1.46	3.41	1.00
SD	2.78	2.44	3.40	1.97
Causation				
Mutable outcome				
M	6.09	2.17	4.84	2.67
SD	2.98	2.81	3.12	2.90
Immutable outcome				
M	2.88	1.23	3.20	0.69
SD	2.29	2.14	3.24	1.38
Compensation				
Mutable outcome				
M	7.15	7.26	4.54	4.79
SD	1.92	2.43	5.58	2.62
Immutable Outcome				
M	5.36	4.20	3.15	2.44
SD	2.75	3.02	2.30	2.53

was seen as more negligent than the nonnegligent one, $F(1, 262) = 211.98$, $p < .0001$. As would be expected, the negligent physician was also blamed more than the nonnegligent physician, $F(1, 262) = 73.74$, $p < .0001$, and seen to have exerted greater causal influence, $F(1, 262) = 115.43$, $p < .0001$.

As in the previous study, significant main effects of outcome mutability were obtained on ratings of blame, $F(1, 262) = 27.56$, $p < .0001$, and causation, $F(1, 262) = 39.73$, $p < .0001$, and on compensation recommendations, $F(1, 262) = 45.19$, $p < .0001$, but not on negligence ratings, $F(1, 262) = 3.99$, $p < .089$. As Table 2 shows, the physician was blamed more, seen as more the cause, and required to pay greater compensation to the patient's family when the patient's death was avoidable versus unavoidable.

The third independent variable, pertaining to the positive or negative victim characterization, yielded significant main effects on compensation recommendations, $F(1, 262) = 55.74$, $p < .0001$, but not on ratings of blameworthiness, $F(1, 262) = 2.40$, $p < .13$, or perceived negligence, $F(1, 262) = 2.15$, $p < .15$.

The main prediction in this study was a three-way interaction among the physician's negligence, outcome mutability, and the victim characterization. This hypothesized interaction was obtained on ratings of blame, $F(1, 262) = 4.44$, $p < .037$, and causation, $F(1, 262) = 3.98$, $p < .05$, although not on compensation recommendations, $F(1, 262) < 1$. Tests of the simple interactions on blame revealed the hypothesized patterns: When the defendant was a likable character whose shooting was inadvertent, the physician was seen as more blameworthy for his demise when the outcome was avoidable versus unavoidable, $F(1, 262) = 6.23$, $p < .014$. This finding corresponds to that in Study 1. When the victim behaved despicably, however, and contributed to the circumstances of his being shot, whether the physician's negligence could have led to a different outcome made less of a difference, thus eradicating the interaction ($F < 1$).

As in the first study, the main effect of negligence on blame remained significant after covarying out causal ratings, $F(1, 261) = 11.98$, $p < .0006$. Because the overall Negligence × Mutability interaction was not significant in this study ($p > .16$), we could not do the corresponding covariance analysis on this interaction pattern.

Study 2 shows, therefore, that negative victim information eliminates the tendency to blame a negligent physician more when the outcome could have been

altered. Because our negative victim characterization not only described the victim as a dislikable character but also described his behavior as contributing to his being shot, we are unable to say definitively whether the reduction in blame to the negligent physician was due specifically to the fact that the victim was a dislikable person or to his causal role in creating his own problems. Regardless of which interpretation is correct, it seems clear that by any standards of Anglo-American jurisprudence or common notions of fairness, a victim's prior actions are irrelevant to assessing the blameworthiness of a person who knows nothing about these actions and negligently harms him.

STUDY 3

We have argued that whether an outcome's mutability influences blame and related judgments depends on the locus of negative evaluations. Mutability makes no difference if an actor's behavior is beyond reproach, but an actor who provides a basis for negative reactions is blamed more when the outcome is avoidable. In the first two studies, the perpetrator's wrongdoing was conveyed in the form of a physician whose negligence led to a patient's avoidable or unavoidable death. Because heightened negligence also leads to perceptions of heightened causal influence, one could argue that causation judgments, rather than negative evaluations of wrongdoing per se, account for our results. This interpretation is countered by the finding that the negligence effects remain significant when causal judgments are removed statistically. Nevertheless, it would be ideal to show that a manipulation that does not logically entail enhanced causation produces the same effects.

In Study 3, rather than assessing negligence, we created a scenario in which an actor who was en route to see his dying mother was delayed for 25 minutes either because he stopped to help an accident victim (positive actor characterization) or because he stopped to make a cocaine deal (negative victim characterization). What varies in this instance is the reason for the actor's delay, which establishes a basis for positive or negative reactions. In either case, however, the delay has the identical causal effect of holding him up for 25 minutes. In the mutable outcome condition, the actor's mother dies just as he arrives at the hospital. In the immutable outcome condition, the actor's mother dies as soon as he sets out, and so the delay does not affect the outcome. We assume that the mutability of the outcome will have no effect on blame when there is no basis for negative evaluations, that is, when the actor is characterized positively. When the actor is characterized negatively, however, he should be blamed more when the delay caused him to miss seeing his mother than when the delay was inconsequential.

Method

Participants, scenarios, and procedure. Study 3 was conducted in one large testing session with a total of 91 participants. This study varied the reason for which a man arrived too late to the hospital to see his dying mother (in order to help someone who had been in an accident or in order to make a drug deal) and the mutability of the outcome (whether he would or would not have made it to the hospital in time if not for the delay).

Common information. On the night of April 23, 2004, Jason Strayhorn received a call from his brother telling him that their mother's medical condition had taken a turn for the worse and that she was not expected to live much longer. Strayhorn's brother told him that their mother was still conscious and was asking to see him. Strayhorn's mother had been in the hospital for 3 weeks and her condition had been progressively deteriorating. Strayhorn lived about 30 miles from the hospital and immediately got into his car and began the trip to the hospital.

Desirable actor information. As he was traveling to the hospital, Strayhorn saw a car on the side of the road that appeared to have just been in an accident. He pulled over to see if he could help and found that a young mother who appeared to be in her late 20s was bleeding from the forehead and appeared semiconscious, and her young child, who Strayhorn estimated to be about 5 years old, seemed to have a broken leg. Strayhorn called for help on his cell phone and waited until the police arrived. He then continued on his way, having been held up for about 25 minutes.

Undesirable actor information. As he was traveling to the hospital, Strayhorn decided to stop off at a friend's house to see if his friend was interested in purchasing some cocaine that he expected to score that weekend. His friend was not home yet, so Strayhorn waited for a while. After he made the drug deal with his friend, he continued on his way, having been delayed for about 25 minutes.

Immutable outcome. As it turned out, Strayhorn's mother died just after his brother called and so there was no way he could have made it to the hospital on time.

Mutable outcome. Unfortunately, Strayhorn's mother died just as he got to the hospital.

Response measures. Participants were asked, "How much do you blame Strayhorn for failing to make it to his mother before she died?" "To what extent do you think that Strayhorn was the cause of his failure to get to the hospital on time?" and "Given the little bit you know about Strayhorn from this event, do you have a

TABLE 3: Means and Standard Deviations for Impression, Blameworthiness, and Causation (Study 3)

	Desirable Reason	*Undesirable Reason*
Impression		
Close outcome		
M	8.82	1.21
SD	2.22	1.64
Far outcome		
M	8.83	1.54
SD	2.01	2.68
Blame		
Close outcome		
M	1.54	9.91
SD	2.18	0.29
Far outcome		
M	0.74	6.79
SD	1.91	3.43
Causation		
Close outcome		
M	4.09	9.68
SD	3.71	1.09
Far outcome		
M	1.48	7.33
SD	2.31	3.29

generally favorable or unfavorable impression of him?" As in the previous studies, ratings were made on 11-point scales ranging from 0 (*not at all blameworthy, not at all the cause,* and *very unfavorable*) to 10 (*extremely blameworthy, very much the cause,* and *very favorable*).

Results and Discussion

Table 3 presents means and standard deviations for impression, blame, and causation ratings.

Ratings of their impression of the actor served as a manipulation check in this study. As intended, more favorable impressions were formed of the actor who stopped to help an accident victim than one who stopped to make a cocaine deal, $F(1, 87) = 269.88$, $p < .0001$.

Main effects were obtained for the reason for delay variable on ratings of both blame, $F(1, 87) = 227.38$, $p < .0001$, and causation, $F(1, 87) = 94.98$, $p < .0001$, showing that the actor was blamed more and ascribed more causal influence for failing to arrive on time at the hospital to see his dying mother when he was delayed by a cocaine deal than when he was helping a person who was injured in a car accident.

Main effects were also obtained on ratings of blame, $F(1, 87) = 16.84$, $p < .0001$, and causation, $F(1, 87) = 17.84$, $p < .0001$, for the outcome's mutability, showing that the actor was blamed more and attributed a greater causal role when the outcome was mutable versus immutable.

The predicted interaction on blame was also obtained, $F(1, 87) = 5.84$, $p < .02$. Simple effects test showed, as predicted, that mutability had no effect on blame attributions for the likable victim, $F(1, 87) = 1.41$, $p < .24$, but that the dislikable victim was blamed more when he nearly missed seeing his dying mother than when he was far from achieving this goal, $F(1, 87) = 21.49$, $p < .0001$. The interaction effect for causation ratings was nonsignificant ($F < 1$).

In this study, we statistically removed causal ratings from the main effect of the reason for the delay on blame attributions and for the interaction between the reason for the delay and mutability. Both the main effect, $F(1, 86) = 62.39$, $p < .0001$, and the interaction, $F(1, 86) = 11.35$, $p < .001$, remained significant in this analysis. These findings are unsurprising given that causation was logically equated in the positive and negative victim characterization conditions of this study. Nevertheless, it is important to know that participants' judgments about causation cannot explain the large effect of the positive versus negative actor characterization. Although one could argue that the actor in the positive characterization condition was, in a sense, more constrained by the need to help an accident victim, the time it took him to help was exactly the same as the time it took the negative character to make a drug deal, and therefore, each delay had the same ultimate causal effect of making the actor late to get to the hospital.

GENERAL DISCUSSION

Counterfactual reasoning theories have been applied most frequently to explain the consequences of experiencing outcomes that could easily have turned out better (upward counterfactuals). A prototypic example is a person missing a winning lottery ticket by one number. Intuition and abundant research evidence suggest that the person who just missed the winning ticket would be more upset than the wholesale loser. Of all the predictions that have issued from different counterfactual reasoning perspectives, the most fundamental is the assumption that negative emotions are heightened following an unfortunate event that could have been averted.

Early studies that applied counterfactual reasoning logic to harmful events made ostensibly straightforward predictions: A person who causes harm will be blamed more when it is easy rather than difficult to imagine a more favorable outcome, and a person who incurs harm will receive more sympathy and compensation (if relevant) when the harmful outcome is mutable. These predictions are predicated on the assumption that unfortunate, mutable outcomes elicit stronger negative reactions than do immutable ones and, therefore, produce more antipathy for perpetrators and greater sympathy for victims.

But these simple predictions belie complications that more recent research has begun to address. The most

obvious complexity is the question regarding which of the potentially large number of mutable factors surrounding an event are likely to influence blame attributions. Mandel and Lehman (1996; also Mandel et al., 2005) have updated counterfactual reasoning theories by showing that when people are asked to mutate harmful events, they focus predominantly on how the events could have been prevented rather than on how they were caused. Mandel and Lehman were concerned with the way people reasoned about harmful events rather than with blame per se. However, we believe that our analysis is generally consistent with theirs in that people are likely to view the factors that arouse the most potent negative evaluations to be the most preventable. In our view, negative evaluations are a more fundamental determinant of blame than preventability is. For example, we would predict that acts of extreme beneficence that cause harm (e.g., injuring an accident victim while trying to help her out of a burning car) are less likely to be cited as causes of harmful events than acts of extreme malevolence (e.g., injuring an accident victim while dragging her out of a burning car to steal her money) even if their negation would equally prevent harm. From the culpable control standpoint, negative evaluations that occur in response to malevolent acts, motives, or dispositions give them a special advantage in causal citation and blame (for supporting evidence, see Alicke, 1992).

The culpable control model also agrees with counterfactual reasoning theories that assert the importance of perceived control in causal citation and blame (e.g., Girotto, Legrenzi, & Rizzo, 1991). But the culpable control model assumes that blame attributions are based not only on control in the usual sense of making something happen but also on culpable control, which also entails negative evaluations that occur in response to the participants in the event, their intentions or actions, or the consequences their actions produce. The phrase *culpable control* reflects the assumption that control estimates are in themselves influenced by the desire to blame someone whose behavior, reputation, or social category has aroused negative reactions.

The negligence manipulations that we used in Studies 1 and 2 can be viewed as evidential factors—that is, factors that moral and legal theorists prescribe for assessing moral blame and/or legal responsibility. In fact, varieties of negligence, along with recklessness and intentional harm, constitute the three main legal categories of wrongdoing. The knowledge that a person negligently or intentionally caused harm is perhaps the most pervasive source of negative evaluations in assessing harmful conduct. However, exerting a causal influence rarely suffices for blame. As we showed in these studies, removing the effect of participants' causal judgments did not eradicate the tendency to blame a negligent physician who caused a patient's avoidable death more than a nonnegligent physician. In addition to causing a harmful outcome, blame requires a source of negative evaluations. In the law, for example, harm that is caused unknowingly, accidentally, from compulsion, in self-defense, or out of necessity is typically outside the bounds of criminal prosecution.

In Study 3, we examined the interplay between mutability and an extra-evidential factor, namely, an actor's extraneous laudatory or despicable behavior. Extra-evidential information includes items such as racial classification (Gerbasi, Zuckerman, & Reis, 1977; Kalven & Zeisel, 1966), social attractiveness (e.g., Alicke, 1992, 1994; Landy & Aronson, 1969), physical attractiveness (Mazzella & Feingold, 1994), the nature and severity of the outcome (Alicke & Davis, 1989; Baron & Hershey, 1988), and character or background information (Alicke, Weigold, & Rogers, 1990; Alicke & Yurak, 1995).[2] Despite being legally inadmissible, each of these factors influences ordinary and juridical judgments. In Study 3, we showed that an actor who was delayed in seeing his dying mother either by a drug deal or by helping an accident victim was blamed more in the former case when undoing the delay would have changed the outcome than when the outcome would have remained the same. This study was important for two primary reasons. First, unlike the first two studies, the negative evaluations that the culpable control model posits occurred in a way that could not possibly have been confounded with causal impact: Both delays held the actor up for the same amount of time, and so attributional differences cannot be attributed to logical differences in the actor's causal role. Second, showing that our results apply to extra-evidential factors as well as to evidential factors extends the scope of the culpable control model in explaining mutability effects for harmful outcomes.

Harmful or offensive actions typically occur in a context in which both evidential and extra-evidential factors can potentially influence blame attributions. The next step in culpable control research is to examine in greater detail the interplay between evidential and extra-evidential factors in their influence on blame. We assume that extra-evidential effects do not occur unless a perpetrator has exerted at least a threshold level of volitional (i.e., intent) or causal control. Even the most perfervid racist, for example, will find it difficult to blame a member of a disliked minority group for outcomes over which he exerted no volitional or causal control. Although numerous studies show that victims are sometimes blamed for the harm they incur, in virtually all of these studies, the victim exerts some causal or volitional control over the event. In many studies on acquaintance rape, for example, the victim, by agreeing to the date or by inviting the man into her house, can be construed to have exerted a degree of causal control, which some observers,

unfortunately, also see as a basis for blame. The culpable control model predicts that blame attributions will be heightened only for participants who have unfavorable reactions to the victim's behavior. We assume that extra-evidential sources of negative evaluation, such as social attractiveness or racial classification, work in a compensatory fashion with evidential sources such as intention, causation, and foresight. When the evidence of an actor's blameworthiness is very strong, extra-evidential information should have little effect on blame. Extra-evidential effects should have their greatest influence on blame when the evidence regarding intent, causation, and foresight is relatively ambiguous.

NOTES

1. The culpable control model includes two basic aspects of control: causal control and volitional control. *Causal control* refers to control over the process by which behaviors produce harmful outcomes, whereas *volitional control* refers to the desire to bring about those outcomes and whether the outcomes were actually effected in the anticipated manner. The scenarios used in counterfactual reasoning research typically involve causal rather than volitional control, and that is the aspect of control that we refer to in this article.

2. Character information is generally legally admissible only during the sentencing phase of a trial. We, and Anglo-American jurisprudence, would consider information about a person's dislikableness to be extra-evidential.

Furthermore, outcome information is not always extra-evidential; in civil cases, the amount of harm or loss incurred is a legitimate basis for awarding damages, and in criminal law, the nature of the outcome sometimes determines the definition of the crime, as in attempted versus completed offenses. The outcome mutability manipulations we use in the present studies pertain to whether outcomes that were beyond the actor's causal influence affect blame and control attributions. It is in this sense that we call outcome information *extra-evidential*.

REFERENCES

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*, 368-378.

Alicke, M. D. (1994). Evidential and extra-evidential evaluations of social conduct. *Journal of Social Behavior and Personality*, *9*, 591-615.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556-574.

Alicke, M. D., & Davis, T. L. (1989). The role of *a posteriori* victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology*, *25*, 362-377.

Alicke, M. D., Weigold, M. F., & Rogers, S. L. (1990). Inferring intentions from motive and outcome: Effects on responsibility judgment. *Social Cognition*, *8*, 286-305.

Alicke, M. D., & Yurak, T. J. (1995). Perpetrator personality and judgments of acquaintance rape. *Journal of Applied Social Psychology*, *25*, 1900-1921.

Antaki, C., & Fielding, G. (1981). Research on ordinary explanations. In C. Antaki (Ed.), *The psychology of ordinary explanations of social behaviour* (pp. 27-55). London: Academic Press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569-579.

Boninger, D. S., Gleicher, F., & Strathman, A. (1994). Counterfactual thinking: From what might have been to what may be. *Journal of Personality and Social Psychology*, *67*, 297-307.

Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident: A meta-analysis of the defensive attribution hypothesis. *Psychological Bulletin*, *90*, 496-512.

Gerbasi, K. C., Zuckerman, M., & Reis, H. T. (1977). Justice needs a new blindfold: A review of mock jury research. *Psychological Bulletin*, *84*, 323-345.

Girotto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, *78*, 111-133.

Goldinger, S. D., Kleider, H. M., Azuma, T., & Beike, D. R. (2003). "Blaming the victim" under memory load. *Psychological Science*, *14*, 81-85.

Hart, H. L. A., & Honore, T. (1959). *Causation in the law.* London: Oxford University Press.

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology*, *46*, 44-56.

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, *107*, 65-81.

Hilton, D. J. (1991). A conversational model of causal explanation. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 51-81). New York: John Wiley.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*, 75-88.

Johnson, J. T. (1986). The knowledge of what might have been: Affective and attributional consequences of near outcomes. *Personality and Social Psychology Bulletin*, *12*, 51-62.

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.), *Judgments under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.

Kalven, H., Jr., & Zeisel, H. (1966). *The American jury.* Chicago: University of Chicago Press.

Landy, D., & Aronson, E. (1969). The influence of the character of the criminal and his victim on the decisions of simulated jurors. *Journal of Experimental Social Psychology*, *5*, 141-152.

Mandel, D. R., Hilton, D. J., & Catellani, P. (Eds.). (2005). *The psychology of counterfactual thinking.* New York: Routledge.

Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology*, *71*, 450-463.

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology*, *24*, 1315-1344.

Miller, D. T., & McFarland, C. (1986). Counterfactual thinking and victim compensation: A test of norm theory. *Personality and Social Psychology Bulletin*, *12*, 513-519.

Miller, D. T., Turnbull, W., & McFarland, C. (1989). Counterfactual thinking and social perception: Thinking about what might have been. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 305-331). New York: Academic Press.

Pyszczynski, T. A., & Greenberg, J. (1981). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology*, *40*, 31-38.

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

Roese, N. J., & Olson, J. M. (Eds.). (1995). *What might have been: The social psychology of counterfactual thinking.* Mahwah, NJ: Lawrence Erlbaum.

Shaver, K. G. (1970). Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology*, *14*, 101-113.

Shaver, K. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness.* New York: Springer-Verlag.

Wong, P. T., & Weiner, B. (1981). When people ask "why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, *40*, 246-259.

Received June 16, 2006
Revision accepted March 17, 2008