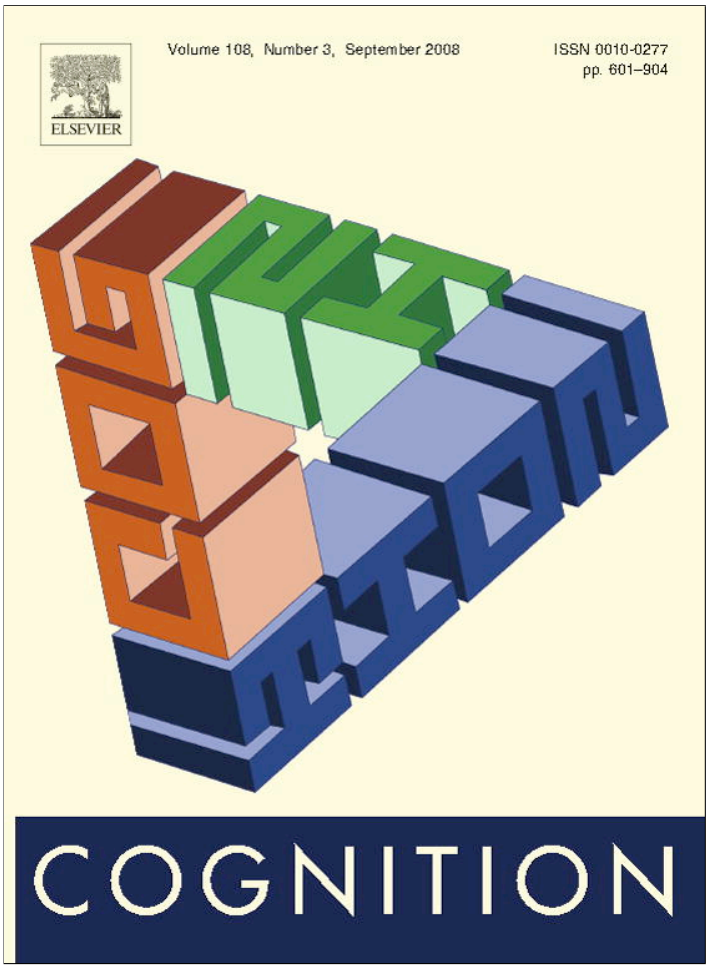

Cognition 108 (2008) 754–770

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Judgments of cause and blame: The effects of intentionality and foreseeability

David A. Lagnado *, Shelley Channon

Department of Cognitive, Perceptual and Brain Sciences, University College London, Gower Street, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 18 September 2007
Revised 8 April 2008
Accepted 29 June 2008

Keywords:
Cause
Blame
Attribution
Intentionality
Foreseeability

ABSTRACT

What are the factors that influence everyday attributions of cause and blame? The current studies focus on sequences of events that lead to adverse outcomes, and examine people's cause and blame ratings for key events in these sequences. Experiment 1 manipulated the intentional status of candidate causes and their location in a causal chain. Participants rated intentional actions as more causal, and more blameworthy, than unintentional actions or physical events. There was also an overall effect of location, with later events assigned higher ratings than earlier events. Experiment 2 manipulated both intentionality and foreseeability. The preference for intentional actions was replicated, and there was a strong influence of foreseeability: actions were rated as more causal and more blameworthy when they were highly foreseeable. These findings are interpreted within two prominent theories of blame, [Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag] and [Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*, 556–574]. Overall, it is argued that the data are more consistent with Alicke's model of culpable control.

1. Introduction

Causal reasoning is fundamental to our ability to predict and control our physical and social environment. It guides our understanding of both the general, the laws that govern events, and of the particular, how and why a specific event actually happened. In everyday life, and in domains such as legal reasoning, the question of actual cause is crucial. Is the accused guilty of causing the victim's death? Is the firm guilty of negligence towards their workers? Such judgments are critical both to the individual and to society at large. An individual seeking to control their environment needs to establish which of their actions are responsible for specific effects. Social institutions seeking to control their individuals need to identify responsible agents and reward or punish them accordingly.

1.1. The problem of causal selection

People are driven to seek causal explanations for how and why things happen (Hart & Honoré, 1985; Heider, 1958; Hilton, McClure, & Slugoski, 2005; Kelley, 1967; Sloman & Lagnado, 2004). However, determining causes is complex in practice. This is especially true of interpersonal situations, where the appreciation of different perspectives may be required, and the success or failure of actions may be judged differently according to the motivations of the agents. In most everyday situations there will be numerous events or factors that could be cited as causes of a particular outcome. For instance, a car accident might be caused by the driver, the condition of the car or road, the inattentive pedestrian, and most likely some combination of all these factors. How do we pick out one cause for the purpose of credit or blame? Attribution research in social psychology has explored this question in some detail (for overview see Fosterling, 2001), but overall the findings are mixed and relatively unsystematic, and often restricted

* Corresponding author. Tel.: +44 (0) 20 7679 5389; fax: +44 (0) 20 7436 4276.
E-mail address: d.lagnado@ucl.ac.uk (D.A. Lagnado).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.06.009

to simple either-or choices rather than more complex causal models (cf. Hilton et al., 2005; Kelley, 1983). In addition, the influence of mental factors such as intentionality and foreseeability have been discussed theoretically (Heider, 1958; Shaver, 1985) but not systematically manipulated in experimental studies (Malle, 2004).

This paper is concerned with how people select causes from chains or networks of events involving human agents. These provide a more realistic model of everyday situations than scenarios that assume only a single preceding event. Chains are ubiquitous in everyday reasoning, especially when one seeks to identify the causes of a significant outcome. Was it lack of sleep that led to the driver's slow reactions, and hence to the collision with the bus? Chains are particularly important in legal reasoning, where a central aim is to trace a causal path from the accused to the crime (Hart & Honoré, 1985; Moore, 1999). Crucial decisions often hang on whether a chain is interrupted or superseded by another's actions. For example, an oil company that negligently allows a tanker to leak petrol is not held responsible in law for a subsequent fire started by an arsonist throwing a cigar onto the petrol (Hart & Honoré, 1985). Attributions in causal chains have received some attention in the literature (Mandel, 2003; Miller & Gunasegaram, 1990; N'gbala and Branscombe, 1995; Spellman, 1997; Vinokur & Ajzen, 1982), but the role of factors such as intentionality and foreseeability have not been explored in any depth.

1.2. Previous research

Theoretical research in social psychology has mainly focused on blame attributions. We will outline two of the most prominent theories, Shaver (1985) and Alicke (2000), with particular emphasis on how they relate to causal judgments. These theories will serve as overarching frameworks to understand and interpret our own experiments.

1.3. Shaver's theory of blame

Shaver (1985) provides a comprehensive theory of responsibility and blame. He builds on the pioneering work of Heider (1958) and Kelley (1973) in social psychology, but also draws on work by legal theorists (Hart & Honoré, 1985) and philosophers (e.g., Austin, 1961; Collingwood, 1940). He advances a prescriptive theory of blame; that is, an account of how an ideal observer ought to make judgments of responsibility and blame. Central to his framework are five dimensions of responsibility (corresponding to the five steps that an observer should consider prior to attributing responsibility to an actor): (1) Causality – the actor's causal contribution to the production of the effect; (2) Knowledge – the actor's awareness of the consequences of their action; (3) Intentionality – the actor's intention to bring about the event in question; (4) Lack of Coercion – whether the actor was coerced into the action; (5) Appreciation of the moral wrongfulness of the action.

As the extent of each of these dimensions increases, so the observer should increase their attribution of responsibility to the actor. However, for the actor to be assigned blame rather than just moral responsibility for their action, two further conditions must be met. The observer should consider any justifications or excuses offered by the actor for their conduct. Only if these are rejected should the actor be blamed for his actions.

An important component of Shaver's theory is his analysis of the dimension of causality. Strongly influenced by Heider (1958), he describes this in terms of a 'causality staircase', climbing up four levels that increasingly connect the actor to the outcome. First is the *association* between the actor and the outcome. This is the weakest level, but can suffice in cases where restitution for the harmful consequences of an action is required. For example, a parent might be held responsible for the havoc wrecked by their unruly child. Second is *causality* itself; for Shaver this is construed as a proximate and generative relation between action and outcome. Third is the *foreseeability* of the outcome by the observer. Shaver argues, in line with Heider, that the more an actor anticipates the negative consequences of their action, the more they should be considered a cause of that outcome. This is explained by the augmentation principle (Kelley, 1973), whereby the causal force of an action is augmented if it has to overcome an obstacle. In this case the obstacle is the actor's foresight about the negative consequences of his action (combined with the affirmative obligation to avoid harm). In short, an actor should be judged as more of a cause when they foresee the adverse consequences of their action, because they have overcome this obstacle. Fourth is the *intentionality* of the actor. This also serves to increase assessments of cause. Thus an intentional action is considered more causal than an unintentional, even if they lead to the same consequence. This is due both to augmentation, where the intentional action must overcome the obstacle provided by the obligation not to cause harm, and because the observer's knowledge that an action is intended often helps to exclude other possible causes of the outcome.

On Shaver's theory intention and foresight play a double role; they are separate dimensions for attributions of responsibility, and are both sub-components in the causality dimension too. However, he emphasizes that they operate in a distinct fashion depending on whether they are serving judgments of causality or of moral responsibility. In the former case, their influence is relatively marginal; most of the weight is carried by the causality component. By contrast, in the latter case, they are both critical to assessments of responsibility.

1.4. Potential shortcomings with Shaver's theory

Shaver's theory is impressive in its scope, and will provide one framework for evaluating the results of our current experiments. However, it is important to note several potential shortcomings with this approach, in particular with its analysis of causality. One concern is the presence of causality itself as a sub-component of the causality dimension. The only way to avoid circularity here is to operate with two different notions of causality (akin to Heider's distinction between personal and impersonal causality). But then it is natural to enquire whether the

personal notion can be dispensed with altogether, and assimilated within the notion of responsibility. A related point is that because the causal dimension is infused with subjective factors such as intention and foresight, it is not clear why it is prescriptive. Why ought one to assign more causality to the same action–outcome pair when the outcome is foreseen rather than unforeseen? Although Shaver's analysis of causation is consistent with some earlier 'action-based' accounts of causation (e.g., Collingwood, 1940; Reid, 1863) it conflicts with most contemporary theories of causation (e.g., Lewis, 1973; Moore, 1999; Pearl, 2000; Woodward, 2003). Addressing such questions, however, lies outside the scope of this paper.

In addition, there are some specific problems with the augmentation principle, and Shaver's use of it to explain why intention and foresight should influence judgments of causality. One concern is how this account applies to extreme cases of evil actions, where the perpetrator has no inclination to avoid negative consequences (i.e., they are not bound by the moral obligation to avoid harm). Does this make their actions less causal than someone who acts despite such an inclination? It is not obvious what should be prescribed in such cases. Furthermore, the augmentation argument only applies to causation of negative events; but why should intentions or foresight not affect causal judgments for positive events too? For example, if a doctor can be a greater cause of a patient's death when he foresees the side-effects of the prescription, why can't he be a greater cause of a patient's survival when he foresees the beneficial effects of his prescription (compared with a doctor who accidentally prescribes the correct drug). The augmentation principle will not apply in such cases, because the causal efficacy of the doctor's actions is not increased by his overcoming the 'obstacle' of foresight. Note that this objection would not be undermined by the finding that in fact people only augment causality in the negative cases. This is because Shaver's account is a prescriptive rather than a descriptive theory – it purports to tell us what people ought to do, not what they actually do. This leads to a more general problem. Shaver offers a prescriptive account, but there is no guarantee that it is also descriptive. This has to be established through extensive experimentation. Indeed, it is likely that people's attributions deviate from those of the ideal observer, due to prevalent cognitive and motivational biases (Alicke, 2000).

1.5. Alicke's culpable control model

Alicke (2000) advances the culpable control model (CCM) to describe the psychological processes that occur when people make ordinary evaluations of responsibility and blame. The model is based on two central assumptions: (1) that people assess potentially blameworthy actions in terms of the actor's personal control over the harmful consequences; and (2) that people make spontaneous evaluations of these actions that encourage blame rather than mitigation. Furthermore, these spontaneous evaluations can have both direct and indirect effects on blame judgments, and on judgments of causality.

1.5.1. Personal control

Personal control reflects the ability to achieve desired behaviours and outcomes or to avoid undesired ones (Fischer, 1986). It is reduced when these options are restricted or excluded. Alicke identifies three kinds of personal control, all of which are critical to evaluations of blame.

1.5.1.1. Volitional behaviour control. The link between mental states and behaviour: whether someone's actions are freely chosen or compelled. This dimension of control will depend on the extent to which an actor's behaviour is purposeful or accidental, and the extent to which he knew what he was doing.

1.5.1.2. Causal control. The link between behaviour and consequences in the world: whether someone's behaviour causes these consequences. This is defined in terms of the actor's causal impact on the outcome in question, and depends on the uniqueness and/or sufficiency of the actor's contribution, the proximity of the actions to the final outcome, and the likelihood that the outcome would have occurred in the absence of the actor's intervention. Note that this dimension of control has no explicit reference to the mental states of the actor, and thus from a prescriptive viewpoint corresponds to a standard theory of physical causation. Indeed, recent theories offered by Pearl (2000) and Woodward (2003), which are founded on the notion of a cause as a control variable, would be ideally suited to fill out this dimension.

1.5.1.3. Volitional outcome control. The link between mental states and the world: whether someone desired and anticipated the consequences. Clearly this dimension of control depends heavily on the two previous dimensions. Someone's control of an outcome requires that they exert sufficient control over their behaviour, and that this behaviour effectively controls the outcome itself. Although this dimension of control is diminished by lack of desire or foresight, an actor can still be judged to have effective volitional outcome control for a harmful outcome that they neither desire nor expect. This will happen if it is judged that they 'should have' foreseen the harmful outcome, and corresponds to the legal notion of 'reasonably' foreseeable.

Alicke's model thus integrates the three main factors from attributional research – causality, intentionality and foreseeability – into dimensions of personal control. Those factors that increase personal control (e.g., effective actions that are intended and expected to cause harm) will increase blame attributions, and those that decrease personal control will mitigate blame. He also accentuates the graded nature of judgments of personal control. In opposition to much previous theorizing, he argues that people do not simply dichotomize factors; they assess the degree of intention, foresight and causality. For example, causal control can vary from weak to strong, and foreseeability can vary from a slight inkling to a strong expectation. This also leads to graded assessments of cause and blame, rather than simple all-or-nothing categorizations.

1.5.2. Spontaneous evaluations

The second novel assumption in Alicke's model is that people engage in spontaneous evaluations of all elements of the situation (e.g., mental, behavioural and consequential aspects). These evaluations are less deliberative than judgments of personal control, and can lead to significant biases in the processing of relevant information. In particular, they typically result in greater blame being ascribed to human agents, and less notice taken of mitigating circumstances.

Spontaneous evaluations are affective reactions to the participators and the harmful events that are caused. They are triggered both by evidential aspects, such as an actor's intentions and foreknowledge, and extra-evidential aspects, such as social attractiveness, race and gender. These evaluations can have both direct and indirect influences on people's subsequent attributions of cause and blame.

A direct influence occurs when an observer responds to the negative consequences of someone's actions (e.g., the distressing nature of the outcome), and therefore attributes blame to the actor, irrespective of careful consideration of the actor's intentions or foresight. For example, when a car driver accidentally kills a child, observers might assign blame to the driver based on an affective reaction to the tragic outcome. This attribution short-circuits the more deliberative route of blame assignment, which would pass through assessments of the driver's personal control.

Alicke, Davis, and Pezzo (1994) explored an extension of this effect, whereby spontaneous negative evaluations led to greater ascriptions of blame, which in turn led to distorted causal control assessments to justify this inflated blame attributions. For example, the car driver might be assigned a greater causal role in the accident in order to justify the observer's spontaneous assignment of blame. This highlights the possibility that, in addition to causal judgments being precursors to blame judgments (Heider, 1958; Shaver, 1985), the link can flow in the opposite direction, with spontaneously arrived at blame ascriptions influencing and distorting causal attributions.

Spontaneous negative evaluations can also influence blame judgments in an indirect fashion. In this case negative evaluations can alter the way in which an observer adjudges aspects of personal control, and this in turn affects the final blame attributions. For example, the observer in the car accident might first exaggerate how much control the car driver had over the accident, perhaps by assuming that he could have foreseen the harmful outcome. This is then used as evidence to support an increased attribution of blame.

In sum, spontaneous evaluations can have strong distorting effects on both blame and cause judgments. Not only do they affect blame judgments directly, but also indirectly, via the modification of assessments of causal or volitional control. There are therefore several pathways by which blame and cause can be linked in psychological processing, some of which deviate from the prescriptive model advanced by Shaver.

1.5.3. Blame-validation & human agency control

The final component of Alicke's model is the assumption that people engage in blame-validation processing. This is the general tendency to assign blame for harmful outcomes and to downplay mitigating circumstances. Spontaneous affective evaluations are one way that this tendency manifests itself. Another arises from the propensity to perceive people rather than the environment as the primary controlling forces underlying negative events (Jones, 1990). As a consequence, evidence that supports an explanation of a harmful event in terms of human agency will be emphasized, at the expense of purely physical explanations that mitigate blame. Alicke points to several psychological reasons for this: human actions seem more controllable than environmental events; they are easier to imagine rectifying (Kahneman & Miller, 1986), and they are often the abnormal feature that makes a difference to the ordinary course of affairs (Hart & Honoré, 1959).

1.6. Intentionality and foreseeability

The two key factors that emerge from Shaver's and Alicke's theories of blame are intentionality and foreseeability. Both play crucial roles in the assignment of blame for negative outcomes. However, there are some important differences between their two accounts. Shaver offers a prescriptive account of blame in which intentions and foresight play a double role: they influence causal judgments (through augmentation) and they separately influence blame judgments. In contrast, Alicke offers a descriptive account. While he agrees that intentions and foresight affect judgments of blame (for him via perceptions of personal control), he also claims that cause and blame attributions are distorted by spontaneous evaluations. In short, for Shaver the influence of intentionality and foreseeability on causal judgments is perfectly rational (what is expected from an ideal observer), whereas for Alicke they represent distortions due to biased processing.

Although Shaver and Alicke cite various experimental studies in support of their respective theories, neither present data explicitly designed to test out their claims. The current experiments will systematically vary intentionality and foreseeability, and will measure judgments of both cause and blame. This should allow us to assess the feasibility of the two accounts.

1.6.1. Intentionality

As noted already, intentionality is a central factor is most attributional theories, including those of Shaver and Alicke. Despite this, there has been little explicit research on the influence of intentionality on people's attributions. One recent exception is Hilton et al. (2005). Based on legal scenarios inspired by Hart and Honoré (1985), Hilton et al. investigated causal attributions in causal chains, and found that people prefer to trace a path back from the outcome (e.g., a car accident) to a human action (e.g., a man flooded the road), but not to an equivalently located physical event (e.g., a storm flooded the road). In contrast to Hart and Honore's dictates their data showed no difference between intentional actions (e.g., the man wished to cause an accident) and unintentional actions (e.g., the man did not consider the consequences). However, this finding was based on only three different scenarios, so it is not clear to what

extent these results generalize. In particular, the failure to find a difference might have been due to idiosyncrasies of the scenarios used. For example, in one scenario the intentional action was 'a man who wished to cause an accident sprayed a road with water' whereas the corresponding unintentional action was 'a man flooded a road without thinking about the consequences'. It is quite possible that the severity of the respective actions (spray with water vs. flood) counteracted the difference in intentionality. To address this problem, in Experiment 1 we use 18 different scenarios, and make sure that the action in question is identical, except for being either intentional or unintentional.

1.6.2. Foreseeability

Foreseeability is another key factor in making attributions. This is supported by a few studies in social psychology (e.g., Fincham & Jaspars, 1983). However, the notion of foreseeability needs to be unpacked, as does its relationship to intentionality. There are at least three varieties of foreseeability: subjective, objective, and reasonable. *Subjective* foreseeability concerns how likely an event is from the agent's point of view; *objective* foreseeability concerns what is in fact likely, irrespective of what the agent actually expects; *reasonable* foreseeability concerns what is reasonable for the agent to expect (what they should expect, given the information available to them).

To illustrate these three notions, consider a doctor who gives a drug to a patient and the patient subsequently dies from an adverse reaction. The subjective aspect of foreseeability corresponds to whether the doctor expected an adverse reaction from the patient. The objective aspect corresponds to whether an adverse reaction was in fact likely (given the objective state of the world). The reasonable aspect corresponds to whether the doctor should have expected the adverse reaction (e.g., whether he should have researched the issue further before administering the drug). This final notion is clearly of importance in legal situations, but is considerably more complex than the other two (Chockler & Halpern, 2003).

Shaver (1985) also distinguishes between subjective and reasonable foreseeability, but does not mention objective foreseeability. On his account, blame attributions should be affected by what the actor should have known about the consequences of their actions (reasonable foreseeability), not what they actually knew (subjective foreseeability). In contrast, he argues that for causal judgments, it is what an actor actually knows, not what they should have known, that is critical. This is because augmentation can only operate against a known obstacle. Thus an actor's causal efficacy is only augmented if they actually foresaw the impending harm. In Experiment 2 both subjective and objective foreseeability are manipulated. This allows us to explore the effects of foreseeability in greater detail, and provides an initial test of Shaver's model.

1.6.3. Interplay between foreseeability and intentionality

Although intentionality and foreseeability are separate notions, there is some interplay between them. For one, it is impossible for an agent to intend an outcome that they do not foresee, and in general the less foreseeable the outcome of an action (from the agent's perspective) the less likely it is to be intentional. Only in specific circumstances will an agent choose an act that they believe to have a very low chance of bringing about the desired outcome (e.g., buying a lottery ticket in order to gain a house). Second, the more foreseeable the outcome of an action, the less likely it is unintended. This lies at the heart of many legal debates. If the accused did something that they knew was highly likely to cause the adverse outcome, it is harder to accept that the outcome was unintended. However, there are clear exceptions. For example, a surgeon who undertakes high risk surgery to save a patient might think death on the operating table likely, but does not thereby intend this outcome. In Experiment 2 both intentionality and foreseeability will be factorially manipulated. Any interplay between them should show up in terms of interactions.

1.7. Location

Another factor implicated in people's attributions is the location of the putative cause in the chain of events leading to the outcome. That is, other things being equal, do people's attributions focus on earlier or later events in the chain? This is hard to establish given that 'other things' are rarely equal, and this perhaps explains the mixed findings in the literature thus far. Several researchers have argued for a primacy effect, where causal attributions are more likely to be directed at the initiating event in a causal chain (Johnson, Ogawa, Delforge, & Early, 1989; Vinokur & Ajzen, 1982). However, arguments have also been made for a recency effect, where attributions are directed to the last event just before the outcome (Einhorn & Hogarth, 1986; Miller & Gunasegaram, 1990; N'gbala & Branscombe, 1995). Likewise, Alicke (2000) suggests that closer proximity between an action and its effect might signal greater causal control over the outcome, and thus a higher degree of causality.

The lack of consistency in these findings can be reconciled to a large extent once different types of chains are distinguished. Drawing upon Hilton et al. (2005), it is useful to distinguish three types of chain. First, *temporal* chains, in which successive events are causally independent. In such cases people's attributions tend to show a recency effect (Miller & Gunasegaram, 1990). Second, *unfolding* causal chains, in which the events are causally dependent, with each successive event following on from and being constrained by the previous one. Here people's attributions tend to show a primacy effect (Miller & Gunasegaram, 1990; Vinokur & Ajzen, 1982), because it is the initial event that sets the causal process in motion. Third, there are what Hilton et al. term *opportunity* chains. In these cases the initial cause creates the opportunity for a later cause to act, but does not necessitate the subsequent cause. Following Hart and Honoré (1985), Hilton et al. argue that in such chains people will show a recency effect, unless the initial cause is a voluntary human action, and the later cause a natural physical event.

An alternative explanation of the mixed findings is given by Spellman (1997). She argued that neither primacy nor recency is more basic, but that attributions depend on the conditional degree of covariation between putative cause and outcome (see below for more details).

As noted above, in their discussions of legal scenarios Hart and Honoré (1985) made the case for an interaction between location and intentionality. They argued that when events are of equivalent status (either two voluntary actions or two physical events), then the most recent event is preferred, but that a voluntary action is always preferred to a physical cause, irrespective of location. McClure, Hilton, and Sutton (2007) supported the latter prediction but failed to find a recency effect with two voluntary actions. Thus, in a chain where one voluntary action (e.g., someone started a small fire) was followed by another (e.g., someone else happened upon the fire, and fanned it to cause a major forest fire), participants rated both actions equally as causes of the blaze. McClure et al.'s conclusions are based on just three scenarios, so it is not clear if these results generalize (especially given the difficulty of controlling for factors other than location). Experiment 1 will focus on opportunity chains, using a wide range of scenarios, and both location and intentionality will be manipulated.

1.8. Judgments of cause vs. blame

When assessing people's attributions, it is important to distinguish between judgments of cause and blame (Chockler & Halpern, 2003; Fincham & Jaspars, 1983; Shaver, 1985; Shultz & Schleifer, 1983). Although closely interrelated, cause and blame are distinct concepts. Someone can cause an outcome, but not be to blame for it; someone can also be blameworthy for an outcome they did not cause. For example, if an infant plays with a loaded gun and shoots someone, they are the cause of the injury, but cannot be blamed for it. In contrast, the parents might be blamed (for leaving a loaded gun in the nursery) even though they did not cause the injury.

As our survey of Shaver and Alicke's theories has shown, the precise relation between cause and blame is complex and controversial. Shaver makes the case for a prescriptive model in which both causal and blame judgments ought to be affected by factors such as intentionality and foreseeability. Alicke's model does not include such a commitment, and implies that any influence of these variables on causal judgments is due to bias rather than normative inference. Moreover, most contemporary theories of causation do not countenance the influence of mental states on assessments of causality.

Fortunately we do not need to get too embroiled in these thorny issues. Our central question is how people's actual causal judgments are modulated by these factors, not whether these attributions fit with a normative theory of causality. Both Shaver and Alicke argue that people's actual causal judgments are likely to be influenced by these factors, albeit for different reasons. More specifically, Shaver would predict only a marginal effect of intentionality and foreseeability (because augmentation is dominated by considerations of causality). Alicke, however, would predict a larger effect, due to the biasing effect of spontaneous evaluations.

1.9. Probabilistic models of causal judgment

Causal judgments about general laws (e.g., smoking causes cancer) are often analyzed in terms of probabilistic models (Cheng, 1997; Shanks, 2004), although the domain is usually restricted to relations between physical events. Probabilistic models have also been applied to the problem of causal selection (Brewer, 1977; Cheng & Novick, 1990; Cheng & Novick, 1992; Spellman, 1997). All of these are based on the degree to which a potential cause covaries with the target effect. The basic claim being that the more the putative cause raises the probability of the effect (from its baseline value), the more likely it is to be selected as the cause of the effect.

Fincham and Jaspars (1983) found some support for Brewer's model with simple models without chains, but did not look at chains with intervening agents. Only Spellman's probability updating account is specifically designed to apply to causal chains. She proposed that people compute a sequence of conditional probability estimates, and select the cause that adds most to the probability of the outcome. However, such a model is blind to the nature of the causes (e.g., action vs. natural physical event) and also to the foreseeability of the outcome from an agent's perspective.

Hilton et al. (2005) and McClure et al. (2007) tested a variety of probabilistic models (including Spellman's), and found that none fitted their data completely. This was largely due to the inability of the models to predict the difference between a human action and a physical event. We expect the same to be true with the current experiments – that the influence of factors such as intentionality and foreseeability cannot be captured by probabilistic models alone. Moreover, unless cause and blame judgments are very similar, it is impossible for a single probabilistic model to capture both kinds of judgment. However, it is still worthwhile assessing the fit of probabilistic models such as Spellman's, because they are the dominant formal models in current theories of causal judgment (Cheng, 1997; Shanks, 2004).

1.10. Overview of experimental paradigm

Experiment 1 used a scenario-based paradigm similar to that employed by Hilton et al. (2005) and McClure et al. (2007). Participants were presented with numerous everyday scenarios, each depicting a chain of events leading to an adverse outcome (see Fig. 1). In the terminology introduced by Hilton et al. these causal chains were opportunity chains (see above). Participants were asked to make cause and blame judgments for two target events in this chain. They also made probability judgments about the likelihood of these events. The key experimental manipulations were the nature of the target events (intentional, unintentional, and physical) and their location in the chain (early, late).

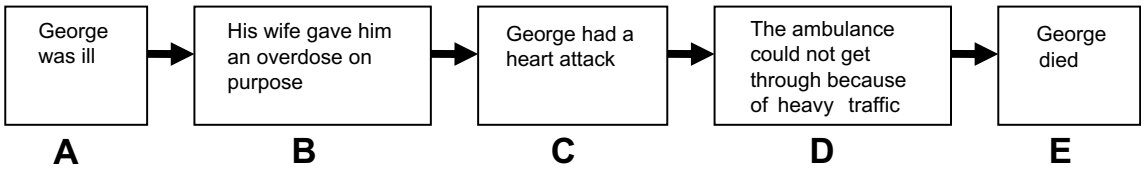

Fig. 1. Simplified example of causal chain used in Experiment 1.

Table 1
Example scenario for Experiment 1

Nature of event	Early (B)	Late (D)
Intentional	"His wife gave him an overdose on purpose"	"The ambulance centre ignored the call"
Unintentional	"His short-sighted wife gave him an overdose by mistake"	"The ambulance driver got lost"
Physical	"The machine labelled the tablets wrongly"	"The ambulance could not get through because of traffic"

George had been ill for a long time. His wife looked after him at home. She was tired of caring for him and gave him an overdose of his tablets. George had a heart attack. His wife phoned the ambulance. However, there was heavy traffic and the ambulance could not get through. The ambulance did not arrive and George died.

2. Experiment 1

2.1. Method

2.1.1. Participants and apparatus

Eighty undergraduates from UCL took part in the study. The experiment was conducted on individual computers with software specially programmed in Visual Basic.

2.1.2. Materials

Eighteen different scenarios were constructed, each with a chain of events leading to an unpleasant outcome (see example in Fig. 1). All scenarios began with an initial background description (A) that set the scene. There were two critical events in each chain: one that occurred *early* in the chain (B), and one that occurred *late* (D), just before the final outcome (E). The nature of these critical events (intentional, unintentional, or physical) was systematically varied so that each scenario compared 2 of the 3 types of events in both early and late positions. Between these two critical events was an intermediate event (C) that served as a bridge between B and D. The initial background (A), intermediate event (C) and final outcomes (E) were kept constant within each scenario.

This yielded 6 versions of each scenario, with comparisons between: intentional vs. physical, physical vs. intentional, unintentional vs. physical, physical vs. unintentional, intentional vs. unintentional, and unintentional vs. intentional. All participants received 3 scenarios of each pair type. An example of a scenario is given in Table 1, along with the key variants for the early and late events (see Appendix for the full set of scenarios).

2.1.3. Procedure

All participants read the following instructions:

You are going to read a series of short stories. They will appear one at a time on the computer screen. There is no need to remember the story. It will stay on the screen. After you have finished reading each story, click the button marked "OK". After reading each story, you will be asked some questions about ***cause*** and ***blame***. As an example, imagine that a baby played with a loaded gun. The gun went off and injured someone. The baby was the ***cause*** of the injury. The baby was not to ***blame*** for the injury. We will ask you to make decisions about ***cause*** and ***blame*** for a range of stories.

Participants were given a practice example before they started the task proper. They each completed 18 problem scenarios, one at a time (event type pairs were counterbalanced across scenarios and participants). For each problem participants were first presented with the causal scenario (see example in Table 1). See Appendix for the full set of scenarios. Once they had read the story they made separate cause and blame judgments, and then gave a set of four probability judgments.

2.1.3.1. Cause and blame judgments. Participants rated the two critical events separately for both cause and blame. In the cause judgments they rated the extent to which each event was to cause for the outcome (on a scale from 0–100, with 0 = not at all the cause, 100 = completely the cause).[1] Blame judgments were elicited in the same fashion. Participants rated the extent to which each event was to blame for the outcome (on a scale from 0–100, with 0 = not at all to blame, 100 = completely to blame). The order of judgments (cause or blame first) was counterbalanced across participants.

2.1.3.2. Probability ratings. After the cause and blame judgments participants gave four separate conditional probability judgments (always in the following fixed order): (i)

[1] We decided to use ratings rather than dichotomous choices for several reasons. They provide a more sensitive measure, and do not force participants into a binary choice; they are consistent with the measures taken in Hilton et al. (2005) and McClure et al. (2007); they fit with Alicke's claim that attributions are graded rather than dichotomous. Moreover, Hilton, McClure, Sutton, Baroux, and Magarou (2008) used both selection (Experiment 1) and ratings (Experiments 2-4) and found no important differences between the two methods of elicitation. The choice of ratings rather than dichotomous choices in the present scenarios cannot therefore explain any discrepancies with previous findings.

the probability that outcome E would have occurred, given that they just knew about the background condition A, P(E|A); (ii) the probability that outcome E would have occurred, given that they knew about the background condition and the early event B, P(E|A&B); (iii) the probability that outcome E would have occurred, given that they knew about the background condition A, the early event B, and the intermediate event C, P(E|A&B&C); (iv) the probability that outcome E would have occurred, given that they knew about the background condition A, the early event B, the intermediate event C, and the late event D, P(E|A&B&C&D). Each probability judgment was registered by entering a number between 0 and 100, with 0 = would definitely not have occurred, 50 = as likely to occur as not, 100 = definitely would have occurred.

2.1.4. Results

Preliminary analysis revealed no effect of the order in which cause and blame questions were asked (cause first vs. blame first, $F(1,78) = 0.78$, ns), so this factor was dropped from subsequent analyses. The cause and blame judgments were then analysed separately.

2.1.5. Cause judgments

An ANOVA was conducted with event type (intentional, unintentional, physical), and location in chain (early, late) as within-subject factors. This revealed a main effect of event type, $F(2,158) = 35.65$, $p < .001$, with intentional events ($M = 66.80$) rated higher than unintentional events (57.32), $p < .001$, intentional events rated higher than physical events (54.66), $p < .001$, and unintentional events rated marginally higher than physical events, $p = .06$. There was also a main effect of location, $F(1,79) = 7.94$, $p < .01$, with events in the late position (62.05) rated higher than those in the early position (57.13), and an interaction between event type and location, $F(2,158) = 7.32$, $p < .01$.

The mean ratings for each event type in early and late locations are shown in Fig. 2. Paired t-tests indicated that intentional events were rated no differently in the early or late location, $t(79) = 1.28$, ns, that unintentional events were rated marginally higher in the late location, $t(79) = 1.78$, $p = .08$, and that physical events were rated higher in the late location, $t(79) = 3.84$, $p < .001$.

2.1.6. Blame judgments

An ANOVA was conducted on the blame judgments, with event type (intentional, unintentional, physical), and location in chain (early, late) as within-subject factors. There was a main effect of event type, $F(2,158) = 143.54$, $p < .001$, with intentional events (76.32) rated higher than unintentional events (53.04), $p < .001$, intentional events rated higher than physical events (47.91), $p < .001$, and unintentional events rated higher than physical events, $p < .001$. There was also a main effect of location, $F(1,79) = 20.89$, $p < .001$, with events in the late position (62.06) rated higher than those in the early position (56.12), but no interaction between event type and location, $F(2,158) = 0.55$, ns.

The mean ratings for each event type in early and late locations are shown in Fig. 3. Paired t-tests indicated that intentional events were rated higher in the late location, $t(79) = 4.01$, $p < .001$, as were both unintentional events, $t(79) = 3.04$, $p < .01$, and physical events, $t(79) = 4.44$, $p < .001$.

2.1.7. The relation between cause and blame judgments

Spearman rank correlations were computed between the cause and blame judgments. These are shown in Table 2, conducted separately for the three types of event and the two locations. All correlations were significant at the 0.01 level, and the correlations were highest for unintentional and physical events.

2.1.8. Probability ratings

The four conditional probability judgments given for each problem were used to compute two difference scores, one for the early cause, one for the late cause. These corresponded to the degree to which the cause in question raised the probability of the final outcome. For the early cause this was the difference between the probability of the final outcome E, given the background condition A and the early cause B, and the probability of the final outcome E given just the background condition A.

Difference Early = P(E|A&B) − P(E|A).

For the late cause this was the difference between the probability of the final outcome E given the background condition A, the early cause B, the intermediate event C,

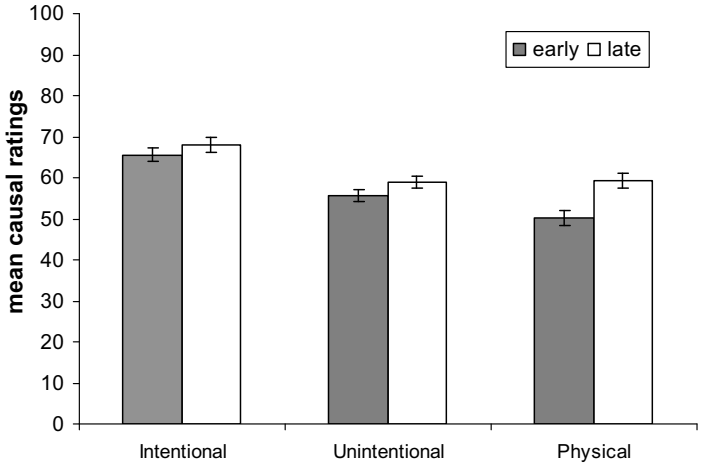

Fig. 2. Mean causal ratings (±SEM) for each event type in early and late locations in Experiment 1.

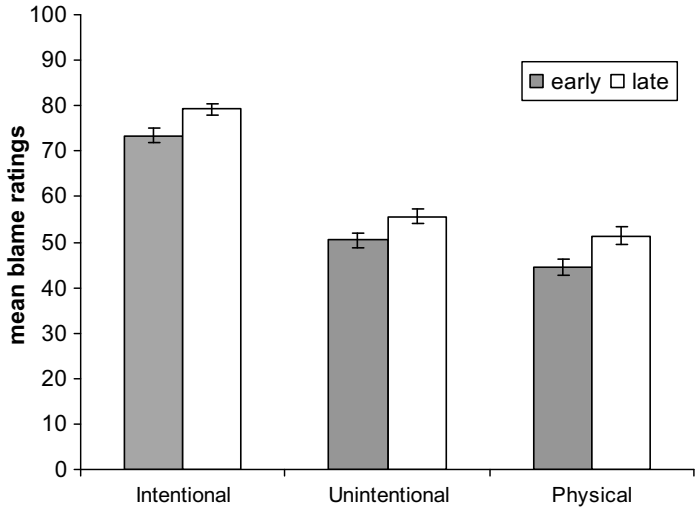

Fig. 3. Mean blame ratings (±SEM) for each event type in early and late locations in Experiment 1.

Table 2
Spearman rank correlations between cause and blame judgments for the three types of event and the two locations

	Early	Late
Intentional	0.336**	0.380**
Unintentional	0.574**	0.415**
Physical	0.578**	0.451**

Note: $^{**}p < .01$.

and the late cause D, and the probability of the final outcome E given all these conditions except the late cause D.

Difference Late = P(E|A&B&C&D) − P(E|A&B&C).

An ANOVA was conducted on the difference scores, with event type (intentional, unintentional, physical), and location in chain (early, late) as within-subject factors. This revealed a main effect of event type, $F(2,158) = 94.22$, $p < .001$, with difference scores for intentional events (38.71) higher than unintentional events (28.41), $p < .001$, and than physical events (25.89), $p < .001$, and unintentional events rated higher than physical events, $p < .01$. There was no effect of location, $F(1,79) = 0.003$, ns, but an interaction between event type and location, $F(2,158) = 11.34$, $p < .001$.

The mean ratings for each event type in early and late locations are shown in Fig. 4. Paired *t*-tests revealed that difference scores for intentional events were higher in the early location, $t(79) = 2.34$, $p < .05$, but for unintentional events there was no difference between locations, $t(79) = 0.14$, ns, and for physical events difference scores were higher in the late location, $t(79) = 2.65$, $p < .05$.

2.2. Discussion

The cause and blame judgments revealed a more complex pattern than that found in previous research. Intentional events were preferred over unintentional and physical events for both cause and blame judgments. This contrasts with the results in Hilton et al. (2005), who found no preference between intentional and unintentional causes, but fits with Hart and Honoré's (1985) claim that voluntary actions are special. This preference for intentional actions is consistent with Shaver's and Alicke's accounts. Both regard intentionality as a key determinant of blame attributions. Moreover, both theories also predict an influence of intentionality on causal judgments. For Shaver this is a prescriptive consequence of his analysis of causality; for Alicke, it is a bias due to people's spontaneous evaluations of negative outcomes, and the consequent exaggeration of causal control.

Blame judgments were also much higher for unintentional rather than physical events. This is consistent with Alicke's claim that people engage in blame-validation, and prefer to assign blame to human agents rather than purely environmental events. In contrast, cause judgments were only marginally higher for unintentional events. This can also be explained on Alicke's account, because blame-validation acts directly on blame judgments, and only indirectly on their causal judgments.

There was a marked effect of location for the blame judgments. For all categories of event more blame was allocated to the later link. There was a more complex pattern for cause judgments. On the one hand, there was no effect of location for intentional events, consistent with the results of McClure et al. (2007). On the other, physical causes in the late position were rated more highly, in contrast to McClure et al. (2007).

As noted in the Introduction, the scenarios in Experiment 1 involved opportunity chains (Hilton et al., 2005), in which earlier causes created the opportunity for later causes to act, but did not necessitate them. Hilton et al. (2005) and Hart and Honoré (1985) argue that people's attributions with opportunity chains will show a recency effect, unless the initial cause is a voluntary human action. Our results fit with these predictions. Moreover, the discrepancy between these judgments and the probability differences scores (which showed a primacy effect for intentional events) rules out Spellman's (1997) claim that the effect of location depends on probability contrasts.

The effect of location on blame judgments fits with Alicke's claim that proximity enhances perceived causal

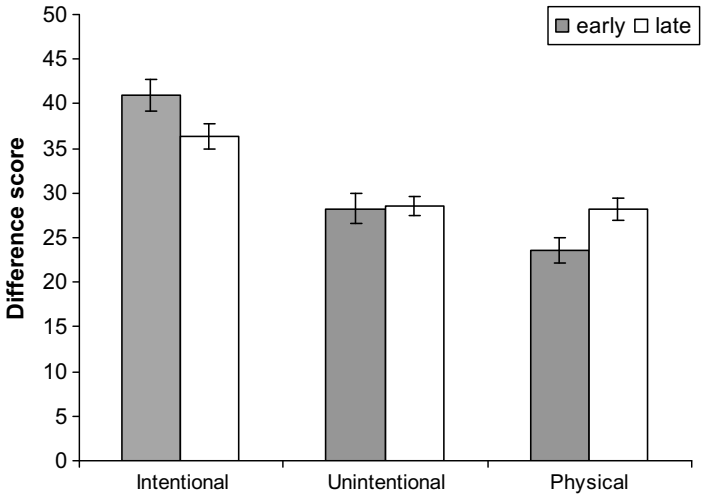

Fig. 4. Mean probability estimate difference scores (±SEM) for each event type in early and late locations in Experiment 1.

control and this in turn heightens blame ascriptions. At first sight, however, this does not explain why, for the causal judgments, there is no significant effect of location with intentional actions. This can be accommodated if we note that on Alicke's account judgments of causal control are responsive to other factors, such as the uniqueness and sufficiency of the action for the outcome. These are commonly measured in terms of the degree to which the cause raises the probability of the effect, and corresponds to the difference scores depicted in Fig. 4. For intentional actions there is a clear preference for the early location, which might have competed with the preference for the late location due to increased perceptions of causal control. Such an explanation would of course require further tests before it could be defended with confidence.

The correlational analyses revealed significant relations between cause and blame judgments in all conditions. This is consistent with both Shaver's and Alicke's theories, which imply a close interrelation between cause and blame judgments.

By taking people's probability ratings we also tested whether a simple probabilistic model could capture people's judgments. Recall that on Spellman's (1997) model people prefer causes that raise the probability of the outcome most. Applying this model to the data in Experiment 1 we found that in general intentional events did raise the probability of the outcome more than unintentional events or physical events, and unintentional events more so than physical events. This pattern actually fits the blame judgments better than the cause judgments. However, the location of events had a different effect for the probability judgments than either cause or blame judgments. In particular, for intentional events there was a preference for the early position which did not correspond to cause or blame judgments.

One complication in interpreting the probability ratings in this experiment (shared with previous studies by Hilton et al., 2005, and McClure et al., 2007) is that these ratings were always taken after the cause and blame judgments. Hence they may have been derived from the prior cause or blame judgments (rather than subserving these judgments). Another concern is that probability ratings were made when participants knew that the outcome in question had in fact occurred, possibly encouraging a hindsight bias (Fischhoff & Beyth, 1975). However, regardless of these complications, the results do suggest that any model simply based on conditional probabilities of events is unlikely to capture the full pattern of human data.

3. Experiment 2

This experiment manipulated both intentionality and foreseeability, and reduced the complexity of the causal networks. In previous research Fincham and Jaspars (1983) focused just on unintentional events, and found that foreseeability influenced both cause and blame ratings, the latter more so than the former. However, the notion of foreseeability was not unpacked, and the relationship between foreseeability and intentionality remained unexplored.

The current study distinguished two kinds of foreseeability: subjective foreseeability, which concerns the degree to which someone expects an outcome, and objective foreseeability, which concerns how likely that outcome really is. We did not use a separate category of 'reasonable' foreseeability, because in the context of the scenarios used in this experiment this overlaps with objective foreseeability.

The accounts of Shaver and Alicke generate differential predictions for these types of foreseeability. As noted above, Shaver claims that judgments of blame hinge on what an actor should reasonably have foreseen, not what they actually foresaw. Thus he would predict that people's blame attributions are sensitive to modulations of objective but not subjective foreseeability. In contrast, for causal judgments Shaver would predict the opposite pattern, because causal augmentation can only occur with actually foreseen outcomes.

On Alicke's account foreseeability modulates the degree of control that an actor is perceived to have over the outcome, and thus has a direct effect on blame attributions. Presumably subjective foreseeability should have a greater

affect on perceptions of personal control, but it is likely that variations in objective foreseeability, coupled with spontaneous evaluations, will also lead to changes in blame attributions. Similar effects would be predicted for causal judgments, but here it is primarily the spontaneous negative evaluations that are doing the work. These evaluations lead people to exaggerate the extent to which the actor has causal control over the outcome.

Finally, as in Experiment 1, both accounts would predict that intentional actions will be attributed more blame and causality than equivalent actions that are unintentional.

3.1. Method

3.1.1. Participants and apparatus

Eighty undergraduates from UCL took part in the study (none had participated in Experiment 1). The experiment was conducted on individual computers with software specially programmed in Visual Basic.

3.1.2. Materials

Twenty scenarios were constructed, 10 with intentional acts and 10 with unintentional acts. Subjective foreseeability (high, low) and objective foreseeability (high, low) were varied such that each scenario had four versions: high subjective and high objective foreseeability (HH), high subjective and low objective foreseeability (HL), low subjective and high objective foreseeability (LH), and low subjective and low objective foreseeability (LL).

Each scenario started with a few sentences that introduced the central agent and set the scene (A). It then stated whether or not the agent expected an outcome from their action (B). After this it stated whether or not the outcome was in fact likely to follow from the action (C). Finally it stated that the outcome had indeed occurred, and noted an adverse consequence of this outcome (D).

Four variants of each scenario were constructed by varying B and C, but keeping A and D constant. A sample unintentional scenario is given (version HH), and all four variations of B and C (HH, HL, LH, LL) are shown in Table 3. See Appendix for the full set of scenarios. Note that in the outcome for all scenarios, the object of the high/low foreseeability manipulation (e.g., the chair collapsing) always occurred, and always led to a final adverse consequence (e.g., the sister is injured).

3.1.3. Procedure

The instructions to participants were very similar to those in Experiment 1. They were told that they would read a series of short stories, and asked to make judgments of cause and blame. The distinction between cause and blame was also illustrated.

Participants were given a practice example before they started the task proper. They each completed 20 scenarios, half with intentional and half unintentional acts. For each scenario participants were first presented with the causal scenario (see example in Table 3). Once they had read the story they made separate cause and blame judgments.

3.1.1. Cause and blame judgments

Participants rated the agent's action for both cause and blame. In the cause judgments they rated the extent to which the agent's action was to cause for the outcome (on a scale from 1 to 4, with 1 = very weakly, 2 = weakly, 3 = strongly, 4 = very strongly). Blame judgments were elicited in the same fashion. Participants rated the extent to which the agent's action was to blame for the outcome (on a scale from 1 to 4, with 1 = very weakly, 2 = weakly, 3 = strongly, 4 = very strongly). Both cause and blame ratings were made on the same screen.

3.2. Results

3.2.1. Causal judgments

An ANOVA was conducted with intention (intentional, unintentional), subjective probability (high, low) and objective probability (high, low) as within-subject factors. This revealed a main effect of intention, $F(1,19) = 366.29$, $p < .001$, with intentional actions ($M = 3.46$) rated higher than unintentional actions (2.57); a main effect of subjective probability, $F(1,19) = 17.40$, $p < .01$, with actions rated higher when the agent expected the outcome (3.08) than when they did not (2.95); and a main effect of objective probability, $F(1,19) = 88.25$, $p < .001$, with actions rated higher when the outcome was objectively likely (3.16) than when it was not (2.87). There were no two-way interactions, but there was a three-way interaction between intention, subjective and objective probability, $F(1,19) = 10.45$, $p < .01$.

Separate ANOVAs were computed for intentional and unintentional actions. For intentional actions there was a

Table 3
Example scenario for Experiment 2 showing the four variants of subjective and objective foreseeability

Foreseeability	Objective HIGH	Objective LOW
Subjective HIGH	Lucy thinks she has not made the chair properly and it is likely to break. Indeed, she has not made the chair properly and it is likely to break (HH)	Lucy thinks she has not made the chair properly and it is likely to break. However, she has made the chair properly and it is unlikely to break (HL)
Subjective LOW	Lucy thinks she has made the chair properly and it is unlikely to break. However, she has not made the chair properly and it is likely to break (LH)	Lucy thinks she has made the chair properly and it is unlikely to break. Indeed, she has made the chair properly and it is unlikely to break (LL)

Lucy buys a self-assembly chair and starts to put it together. The instructions are very confusing. Lucy thinks she has not made the chair properly and it is likely to break. Indeed, she has not made the chair properly and it is likely to break. Lucy's sister comes into the room and sits down on the new chair. The chair collapses beneath Lucy's sister, injuring her.

main effect of subjective probability, $F(1,19) = 24.00$, $p < .001$, with actions rated higher when the agent expected the outcome (3.52) than when they did not (3.40), and a main effect of objective probability, $F(1,19) = 42.67$, $p < .001$, with actions rated higher when the outcome was objectively likely (3.58) than when it was not (3.34). These data are shown in Fig. 5. There was also an interaction between objective and subjective probability, $F(1,19) = 10.51$, $p < .01$.

For the unintentional actions there was a marginal effect of subjective probability, $F(1,19) = 4.01$, $p = .06$, with actions rated higher when the agent expected the outcome (2.63) than when they did not (2.51), and a main effect of objective probability, $F(1,19) = 37.42$, $p < .001$, with actions rated higher when the outcome was objectively likely (2.75) than when it was not (2.39). These data are also shown in Fig. 5. There was no interaction between objective and subjective probability, $F(1,19) = 2.44$, ns.

3.2.2. Blame judgments

An ANOVA was conducted with intention (intentional, unintentional), subjective probability (high, low) and objective probability (high, low) as within-subject factors. This revealed a main effect of intention, $F(1,19) = 956.05$, $p < .001$, with intentional actions (3.45) rated higher than unintentional events (2.41); a main effect of subjective probability, $F(1,19) = 134.85$, $p < .001$, with actions rated higher when the agent expected the outcome (3.12) than when they did not (2.74); and a main effect of objective probability, $F(1,19) = 49.18$, $p < .001$, with actions rated higher when the outcome was objectively likely (3.05) than when it was not (2.81). There were no two-way or three-way interactions.

Separate ANOVAs were computed for intentional and unintentional actions. For intentional actions there was a main effect of subjective probability, $F(1,19) = 51.95$, $p < .001$, with actions rated higher when the agent expected the outcome (3.61) than when they did not (3.30), and a main effect of objective probability, $F(1,19) = 21.05$, $p < .001$, with actions rated higher when the outcome was objectively likely (3.54) than when it was not (3.37). These data are shown in Fig. 6. There was no interaction between objective and subjective probability, $F(1,19) = 0.69$, ns.

For the unintentional actions there was a main effect of subjective probability, $F(1,19) = 63.53$, $p < .001$, with actions rated higher when the agent expected the outcome (2.64) than when they did not (2.18), and a main effect of objective probability, $F(1,19) = 30.84$, $p < .001$, with actions rated higher when the outcome was objectively likely (2.56) than when it was not (2.26). These data are also shown in Fig. 6. There was no interaction between objective and subjective probability, $F(1,19) = 0.69$, ns.

3.2.3. The relation between cause and blame judgments

Spearman rank correlations were computed between the cause and blame judgments. Overall there was a relatively strong correlation, Spearman's rho = 0.56, $p < .01$. Correlations were also computed for intentional and unintentional actions separately. The strength of correlation was slightly more for intentional events, Spearman's rho = 0.49, than for unintentional events, Spearman's rho = 0.36, but both were still significant, $p < .01$.

3.3. Discussion

As in Experiment 1 intentional actions were rated more highly than unintentional actions, for both causal and blame judgments. This is consistent with both Shaver and Alicke. As noted, Shaver would claim that intentionality is a central dimension for blame attributions, and an influence on cause judgments via augmentation. For Alicke, increased intentionality leads to greater perceptions of volitional and causal control.

With regard to foreseeability, the blame judgments showed a more clear-cut pattern than the causal judgments.

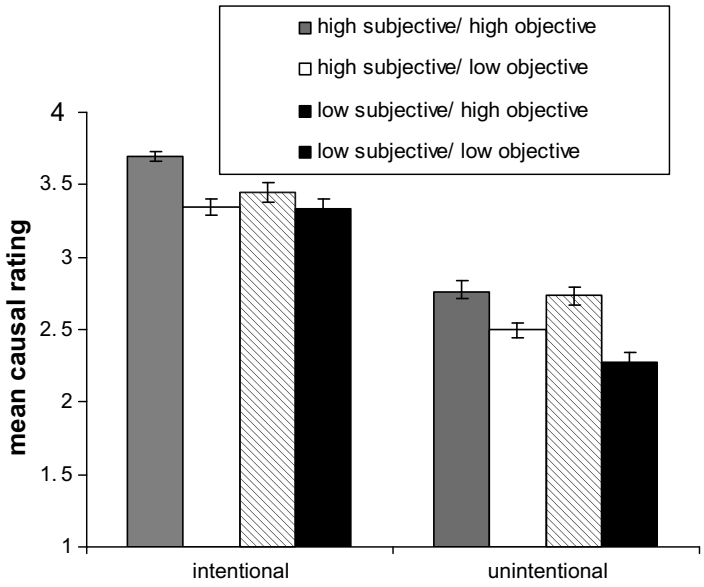

Fig. 5. Mean causal ratings (±SEM) as a function of intentionality and subjective and objective foreseeability.

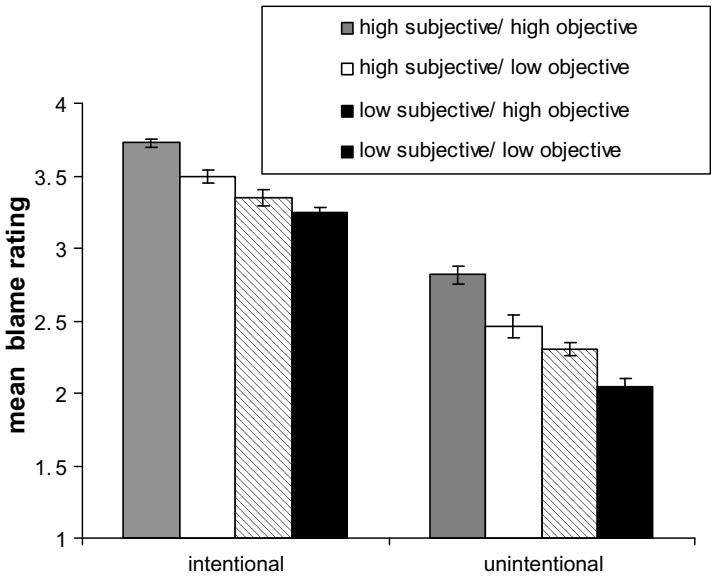

Fig. 6. Mean blame ratings (±SEM) as a function of intentionality and subjective and objective foreseeability in Experiment 2.

Both subjective and objective foreseeability had a strong influence on people's judgments of blame. They rated an agent as more blameworthy when the agent expected the adverse outcome, and when the outcome was indeed more likely. This pattern was unaffected by whether the actions were intended or unintended.

This pattern is more consistent with Alicke's than Shaver's account. Shaver's account implies that blame attributions are affected by changes to what it is reasonable to foresee (in this case objective probability), but not what is actually foreseen (subjective probability). However, both factors had strong effects on blame judgments. This fits better with Alicke's account, which predicts that both factors will influence judgments of personal control.

The effect of foreseeability on the causal ratings was more complicated. Objective foreseeability had a large effect on causal judgments for both intended and unintended actions: the more likely an outcome actually was the more the action was judged as a cause.[2] Subjective foreseeability, however, only influenced causal judgments for: (a) intentional actions that were objectively likely to produce the outcomes, and (b) unintentional actions that were objectively unlikely to produce the outcomes. In both these cases participants judged an action as more causal when the agent thought the adverse outcome likely. In contrast, subjective foreseeability had little effect on causal judgments with: (c) intentional actions that were objectively unlikely to produce the outcomes, and (d) unintentional actions that were objectively likely to produce the outcomes.

The overall pattern of these results is complex, and the exact details are not readily explained by either Shaver's or Alicke's model. Despite this, the robust effect of objective foreseeability is more consistent with Alicke than Shaver. This is because Shaver holds that causal augmentation only occurs when outcomes are actually foreseen (subjective foreseeability), not when they are merely foreseeable. Thus his model implies no independent effect of objective foreseeability on cause judgments. In contrast Alicke's model makes no firm differential predictions, and thus accommodates effects of both objective and subjective foreseeability.

One potential explanation for the complex pattern is that cases (c) and (d) are atypical. Except in special circumstances, intentional actions tend not to be paired with highly unlikely consequences, and unintentional actions tend not to be paired with highly likely consequences. Given that we did not tailor the scenarios to support these less plausible cases, it is possible that participants had a general difficulty in answering causal questions here. Clearly further research is needed to establish this, or explore alternative accounts for this pattern of results.

4. General discussion

Two experiments explored several factors that influence everyday attributions of cause and blame. One major finding was that intentional actions were rated as more causal, and more blameworthy, than unintentional actions. This fits with the theories of blame advanced by Shaver (1985) and Alicke (2000), but contrasts with recent work reported in Hilton et al. (2005). A second finding was that the foreseeability of outcomes also influenced both cause and blame judgments. The more that the effects of some-

[2] It is possible that in some scenarios participants in the low objective foreseeability condition rated the target action as low in cause and blame because they thought that additional, unspecified causal factors played a more significant role. This potentially applies to any situation illustrating low objective foreseeability. For instance, consider the example in the Introduction of a doctor giving a drug with an unknown side-effect leading to an adverse outcome that could not have been predicted (i.e. objectively unforeseeable). It is possible that people might invoke other, unspecified causes in addition to the drug as contributory factors, perhaps playing a larger causal role than the drug itself. Nevertheless, there remains a difference between what the protagonist should have known in the low objective foreseeability condition (e.g. that the drug was not likely to cause an adverse outcome), as compared with in the high objective foreseeability condition (e.g. that the drug was likely to cause an adverse outcome).

one's actions were expected (whether by the agent themselves, or from an objective viewpoint) the more causal and blameworthy they were judged. This fits better with Alicke's model than with Shaver's. This is because Shaver predicts that causal attributions to an actor should be affected by what the actor actually knows, not what they should have known; in contrast, he predicts that blame attributions should be most affected by what an actor should have known, not what they actually know. Neither prediction is borne out by the data. Third, we found an effect of location on people's judgments. More blame was allocated to a later rather than an earlier event, irrespective of whether the events were intentional, unintentional or purely physical. In contrast, for causal judgments the location only mattered for purely physical events (where again later events were rated higher). These findings can also be accommodated on Alicke's model.

Overall, our data provide better support for Alicke's model rather than Shaver's. Moreover, when both models agree, Alicke's model seems to provide a more parsimonious and plausible account. For example, Alicke explains the influence of intentionality on causal judgments by assuming that people exaggerate the degree to which an actor has causal control over a negative outcome. Shaver requires the more problematic notion of augmentation, and needs to argue that this is in fact a normative inference to make.

There are several additional reasons that favour Alicke's position. Alicke presents a psychological model that takes into account the various cognitive and affective biases known to beset human reasoners. By contrast, Shaver's is a prescriptive model, and is supposed to concern how people ought to make attributions, not necessarily what they actually do. Only when people approximate ideal observers would Shaver's account yield a descriptive theory; but this seems the exception rather than the rule, especially in the affectively charged situations typical of blame attributions. Moreover, Shaver's model assumes an atypical (by modern standards) notion of causality, where the same action can have a different causal force according to the mental states of the actor performing it. In contrast, Alicke's model is consistent with more standard normative accounts of causality, and indeed resonates with recent theories that define causality in terms of control variables (Pearl, 2000; Woodward, 2003). Although the notions of control and manipulation play a central role in these theories, neither is inherently tied to human agency (Woodward, 2003). In addition, Alicke's culpable control model ties in with recent work in the psychology of causal learning and reasoning that accentuates the significance of interventions (Gopnik et al., 2004; Lagnado & Sloman, 2004; Lagnado & Sloman, 2006; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005). Underpinning this work is the belief that people's core notion of causality emerges from their ability to control and manipulate the world around them.

Another advantage with Alicke's culpable control model is that it is naturally extended to include notions of social or societal control. This allows it to link with recent work that uses a social functionalist framework to explain people's judgments and choices (Tetlock, 2002). For example, Tetlock argues that people often act as 'intuitive prosecutors' aiming to enforce social norms. This incorporates a commitment towards fairness (e.g., impartial weighing of evidence; respect for human rights etc.) but also introduces affect-induced biases similar to those present in Alicke's account (Goldberg, Lerner, & Tetlock, 1999).

Overall our results support Alicke's model, but there are several components of his model that remain untested. Building on the current experimental paradigm, future research could measure people's perceptions about volitional and causal control; manipulate affective responses and hence spontaneous evaluations; investigate both the direct and indirect routes by which spontaneous evaluations are supposed to influence judgments of cause and blame.

4.1. Dual-aspect of causal judgments

From a broader perspective, our results can be understood in terms of the different functional roles that causal judgments play in everyday cognition. Although our data do not speak directly to this issue, we suggest that causal attributions serve two primary functions: (1) Backward-looking: to facilitate the assignment of blame or credit for what has happened in the past, for the purposes of restitution or compensation. This involves tracing a causal path back from the outcome in question to the potential causes of that outcome. (2) Forward-looking: to avoid adverse outcomes in the future.[3] This involves tracing a causal path forward from causes to potential outcomes.

These two functions are interrelated but distinct. The backward-looking aspect is implicated in typical cases of blame attribution. Who was at fault for this specific outcome? Who should be punished? It is also tied up with questions of legal or moral responsibility (Shultz & Schleifer, 1983). The legal system wants to know who to hold responsible, and perhaps punish or reprimand them. This is the aspect of causal attribution that drives people's judgments of blame, and is integrally linked to questions of intentionality and foreseeability.

The forward-looking aspect is closer to the general notion of causes as potential means for control and manipulation (Pearl, 2000; Woodward, 2003, 2008). People seek to identify causes so that they can make appropriate changes in the future; so that they can avoid adverse outcomes, and bring about advantageous ones. This aspect of causal attribution should be independent of the mental states of the actors involved insofar as the judged causal relation between an action and an outcome should not increase in strength just because the outcome was intended rather than unintended, or foreseen rather than unforeseen. However, this does not mean that these factors are irrelevant in the forward-looking case. In situations that involve human agents, information about an actor's intention and foresight will often help pick out those causes that can be

[3] Woodward (2008) also draws this distinction, and a similar distinction is made in legal reasoning between ex ante and ex post (Farnsworth, 2007). Note that the distinction does not preclude the possibility that people make decisions in order to avoid future blame. In such cases they are anticipating how people might trace back a causal path from the outcome to their action. However, this is separate from the forward-looking sense of cause, which involves imagining the possible consequences of the action.

controlled or manipulated in the future. For example, one way to prevent a doctor prescribing the wrong drug in future cases is to improve their foreknowledge of the adverse consequences. Similarly, one way to stop someone repeating a criminal act is to suppress or eliminate their harmful intentions. In both cases, the manipulation of an actor's mental states serves as a means through which future behaviour can be adjusted (hopefully for the better).

This dual-aspect of causal attribution – as both forward and backward-looking – helps explain the interrelation found between judgments of cause and blame in our experiments, and the marked influence of intentionality and foreseeability on both kinds of judgment. In particular, people rate actions carried out with intent and foresight as more causal because they are more blameworthy (backward-looking), and because such actions are candidates for change in the future (forward-looking). It affords a middle ground between Alicke's and Shaver's positions. It allows for a prescriptive element to the effect of intentionality and foreseeability on causal judgments, because these factors are often crucial for causal control in the future. However, it also accepts that people's judgments of causal control can be distorted, especially when the outcomes are harmful and trigger negative evaluations.

4.2. Relation between normative and descriptive theories of causation

In this paper with have been primarily concerned with people's actual judgments, rather than what is mandated by a normative theory. As noted in the Introduction, the relation between normative and descriptive theories of causation is complex and controversial (see White (1990) for extensive coverage of both kinds of theory). Most contemporary normative accounts do not allow for the influence of intentionality or foreseeability on causal judgments. For example, counterfactual theories (Collins, Hall, & Paul, 2004; Lewis, 1973) make no reference to such factors; causal relations are simply defined in terms of the counterfactual dependencies between events. The same holds true of manipulation theories (Pearl, 2000; Woodward, 2003), although, as mentioned above, the analysis of causation in terms of potential control variables is readily adapted to a descriptive theory such as Alicke's.

In contrast, Shaver (1985) follows on from the earlier 'activity' theories of Reid (1863) and Collingwood (1940) and places human agency at the core of a prescriptive theory of causation. We maintain that this move is unnecessary; that a normative theory (even if defined in terms of manipulation and intervention) does not need to build anthropomorphic concepts into the analysis of causation. We lack the space to argue for this position in detail, but an important reason is the belief in the objectivity and mind-independence of causal relations. As stated above, we maintain that a causal relation between action and outcome should not be accorded greater strength just because the action is intentional, or the outcome foreseen. However, we do think that issues arising from human agency (in particular, intentionality and foreseeability) play critical roles in people's everyday assessments of causality. We trace this to both the forward and backward-looking functions of causal judgments, and to the distorting effects of negative evaluations (in accordance with Alicke's theory of culpable control).

4.3. Probabilistic models

How well can a simple probabilistic model account for people's attributions? Experiment 1 tested Spellman's (1997) model, which claims that people prefer causes that increase the probability of the target outcome. In line with Hilton et al. (2005) and McClure et al. (2007) we found that this model was unable to capture the full pattern of human data. In particular, it did not capture the differential effects of location on cause and blame judgments. The results in Experiment 2 are also difficult to accommodate within a simple probabilistic model. This is because people were sensitive to two different types of probability, both objective and subjective (from the agent's viewpoint). However the simple probabilistic model allows for only one kind of probability estimate.

To capture the pattern of results found in our experiments a probabilistic model would need to be extended to: (i) explain the differences between cause and blame judgments (especially with respect to location); (ii) account for the effects of intentionality and foreseeability; and (iii) incorporate estimates of both subjective and objective probability. Chockler and Halpern (2003) provide a formal account of cause and blame (based on structural models) that is a first step in this direction. Their model has the capability to deal with the impact of foreseeability on blame judgments, because it allows for variations in an actor's epistemic states (in terms of subjective probability distributions). However, the model is not readily adapted to cover the influence of intentionality. This is because it treats causes equally, irrespective of whether they involve human agency or are purely physical.

The inadequacies of a purely probabilistic model have been highlighted in previous work on causal attribution (e.g., Hilton et al., 2005; Mandel, 2003; McClure et al., 2007). However, it remains an important benchmark model against which to test other models. Moreover, any comprehensive psychological model of attribution must explain people's sensitivity to probabilistic contrasts, even if this needs to be supplemented to capture their sensitivity to factors such as intentionality and foreseeability.

4.4. Future work

The two experiments in this paper have focussed exclusively on negative outcomes. This is natural if the main concern is with questions of blame; however, a fuller understanding of causal judgments, and the effects of intentionality and foreseeability, would benefit greatly from looking at positive outcomes too. Would there be similar influences of intentions and foresight in such cases? Or would their influence be restricted to adverse outcomes? This would also provide a critical test of the augmentation principle, which predicts that causal attributions to positive outcomes should not be augmented by intentionality or foreseeability (see the discussion of this in the Introduction). In contrast, the finding that these fac-

tors did influence causal assessments for positive outcomes could be accommodated within Alicke's model. Affective evaluations can be positive as well as negative, and so creditworthy outcomes might also distort people's perceptions of how much control the actor had over these outcomes. Presumably highly positive events would increase the perception of the agent's degree of control. However, it is also likely that the distorting effect of blame induced by negative outcomes will be more powerful than that of credit induced by positive outcomes. In addition, if the attribution of blame is placed in a broader social context, in terms of society imposing punishments and sanctions on its members (cf. McClure et al., 2007; Tetlock, 2002), it is quite possible that positive outcomes will not exert as strong a distorting effect on causal judgments as negative outcomes do. These are empirical question that need to be followed up in future work.

Another natural extension of the current studies would be to combine the foreseeability manipulations in Experiment 2 with longer causal chains. This is important because foreseeability might interact with location in the chain. For example, Hilton et al. (2005) showed that if the earlier event in a chain has foreseeable effects on the final outcome, this can modulate perceptions of the penultimate events in the chain.

4.5. Causal explanations and counterfactuals

There is a substantial literature on the connection between causal explanation and counterfactuals. The adequacy of philosophical analyses of causation in terms of counterfactuals has been the subject of much debate (Collins et al., 2004), and more recently psychological studies have shown that causal judgments can be dissociated from counterfactual judgments (Mandel, 2003; McEleney & Byrne, 2006). These studies did not compare cause and blame judgments, so it is not clear to what extent people's causal judgments were influenced by issues of blame. To follow this up, our current experiments could be extended by including appropriate counterfactual questions, and seeing how these map onto cause and blame assessments. One possibility, consistent with the findings by McEleney and Byrne (2006), is that counterfactual judgments serve mainly for backward-looking attributions; causal explanations for forward-looking attributions.

4.6. Reasonable foreseeability

Our experiments have shown that both intentionality and foreseeability influence people's attribution judgments. However, Experiment 2 only looked at subjective and objective foreseeability. As argued in the Introduction, in many real-world situations, and in legal cases in particular, it is likely that the notion of reasonable foreseeability plays an important role. People are not just concerned with what the agent in question actually foresaw, but also with what they could and should have foreseen (recall the doctor who should have checked his patient's notes before administering a drug). Indeed the results in Experiment 2 suggest that people are sensitive to this distinction. This is shown by the fact that participants' blame judgments were independently affected by both objective and subjective foreseeability. The effect of subjective foreseeability translates straightforwardly to the notion that an agent is more blameworthy if they actually expected their action to have an adverse outcome. The effect of objective probability implies that they also consider what the agent could have expected (what was objectively to be expected). Thus an agent is judged more blameworthy if the adverse outcome was in fact to be expected, even if the agent did not actually expect it themselves. Future research should explore this notion of reasonable foreseeability, and how it relates to other work on expectations and accountability (Markham & Tetlock, 2000).

5. Conclusions

The current experiments have established that intentionality and foreseeability exert strong effects on judgments of both cause and blame. These results can be interpreted within Alicke's culpable control model of blame, whereby intention and foresight have a straightforward influence on blame attributions, and a correlated 'distorting' effect on causal judgments. We also suggest that this fits with the notion that causal judgments about adverse outcomes serve two primary functions: backward-looking to assign blame for harmful events in the past, and forward-looking to avoid harmful events in the future.

Acknowledgements

We are grateful to the ESRC (Grant Ref 23-0959) for supporting this work. Thanks also to Helena Drury and Sian Fitzpatrick for help with materials and data collection.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cognition.2008.06.009.

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*, 556–574.
Alicke, M. D., Davis, T. L., & Pezzo, M. V. (1994). A posteriori adjustment of a priori decision criteria. *Social Cognition, 12*, 281–308.
Austin, J. L. (1961). *Philosophical papers*. Oxford: Clarendon Press.
Brewer, M. B. (1977). An information-processing approach to attribution of responsibility. *Journal of Experimental Social Psychology, 13*, 58–69.
Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.
Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545–567.
Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99*, 65–82.
Chockler, H., Halpern, J. Y. (2003). Responsibility and blame: A structural-model approach. In *Proceedings of the 18th international joint conference on artificial intelligence* (pp. 147–153).
Collingwood, R. G. (1940). *An essay on metaphysics*. Oxford: Clarendon Press.
Collins, J. D., Hall, E. J., & Paul, L. A. (Eds.). (2004). *Causation and counterfactuals*. MIT Press.
Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 1–19.

Farnsworth, W. (2007). *The legal analyst*. University of Chicago Press.
Fincham, E. D., & Jaspars, J. M. (1983). A subjective probability approach to responsibility attribution. *British Journal of Social Psychology, 22*, 145–162.
Fischer, J. M. (Ed.). (1986). *Moral responsibility*. Ithaca, NY: Cornell University Press.
Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior and Human Performance, 13*, 1–16.
Fosterling, F. (2001). *Attribution: An introduction to theories, research, and applications*. Hove: Psychology Press.
Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology, 29*, 781–795.
Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*(1), 1–31.
Hart, H. L. A., & Honoré, A. M. (1959/1985). *Causation in the law* (2nd ed.). Oxford University Press.
Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
Hilton, D. J., McClure, J., & Slugoski, B. (2005). Counterfactuals, conditionals and causality: A social psychological perspective. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 44–60). London: Routledge.
Hilton, D. J., McClure, J., Sutton, R., Baroux, A., Magarou, I. (2008). Selecting explanations from unfolding causal chains: Do statistical principles explain preferences for voluntary causes? Unpublished manuscript.
Johnson, J. T., Ogawa, K. H., Delforge, A., & Early, D. (1989). Causal primacy and comparative fault: The effect of position in a causal chain on judgments of legal responsibility. *Personality and Social Psychology Bulletin, 15*, 161–174.
Jones, E. E. (1990). *Interpersonal perception*. New York: Macmillan.
Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.
Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation*. Lincoln: University of Nebraska Press. Lincoln: Lincoln.
Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28*(2), 107–128.
Kelley, H. H. (1983). Perceived causal structures. In J. M. F. Jaspars, F. D. Fincham, & M. R. C. Hewstone (Eds.), *Attribution theory and research* (pp. 343–369). London: Academic Press.
Lagnado, D. A., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 856–876.
Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory & Cognition, 32*, 451–460.
Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.
Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.
Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General, 137*, 419–434.
Markham, K. D., & Tetlock, P. E. (2000). Accountability and close-call counterfactuals: the loser that nearly won and the winner who nearly lost. *Personality and Social Psychology Bulletin, 26*, 1213–1224.
McEleney, A., & Byrne, R. M. J. (2006). Spontaneous counterfactual thoughts and causal explanations. *Thinking & Reasoning, 12*(2), 235–255.
McClure, J., Hilton, D. J., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European Journal of Social Psychology, 37*, 879–901.
Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology, 59*, 1111–1118.
Moore, M. S. (1999). Causation and responsibility. In E. Frankel Paul, F. D. Miller, & J. Paul (Eds.), *Responsibility*. Cambridge: Cambridge University Press.
N'gbala, A., & Branscombe, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology, 31*, 139–162.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.
Reid, T. (1863). Of the liberty of moral agents (6th ed.. In W. Hamilton (Ed.). *The works of Thomas Reid, D.D.* (Vol. 2, pp. 599–636). Edinburgh: Machlachlan & Stewart.
Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.
Shultz, T. R., & Schleifer, M. (1983). Towards a refinement of attribution concepts. In J. Jaspars, F. Fincham, & M. Hewstone (Eds.), *Attribution theory and research* (pp. 37–62). New York: Academic Press.
Shanks, D. R. (2004). Judging covariation and causation. In N. Harvey & D. Koehler (Eds.), *Blackwell handbook of judgment and decision making*. Oxford: Blackwell.
Sloman, S. A., & Lagnado, D. A. (2004). Causal invariance in reasoning and learning. In B. Ross (Ed.). *The psychology of learning and motivation* (Vol. 44, pp. 278–325). San Diego: Elsevier Science.
Sloman, S. A., & Lagnado, D. (2005). Do we "do"? *Cognitive Science, 29*, 5–39.
Spellman, B. (1997). Crediting causality. *Journal of Experimental Psychology: General, 126*, 323–348.
Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice. Intuitive politicians, theologians and prosecutors. *Psychological Review, 109*, 451–471.
Vinokur, A., & Ajzen, I. (1982). Relative importance of prior and immediate events: A causal primacy effect. *Journal of Personality and Social Psychology, 42*, 820–829.
Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 216–227.
White, P. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin, 108*, 3–18.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
Woodward, J. (2008). Psychological studies of causal and counterfactual reasoning. Unpublished manuscript.