JUDGMENTS IN CAUSAL CHAINS:
THE IMPACT OF POSITIVE AND NEGATIVE MOTIVES AND OUTCOMES
ON
LAY ATTRIBUTIONS

BY

BRIAR HELEN MOIR

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy

Victoria University of Wellington

2014

Acknowledgments

I would like to thank the many people who have helped bring this thesis to fruition. Firstly, to my supervisor John McClure; thank you for all your insights, your generosity of time, and your patience as you guided, challenged and provoked my thinking throughout this research. To my secondary supervisors, Matt Crawford and Garth Fletcher, thank you also for your insights and encouragement. Thank you to those who provided extremely useful comments and encouragement, including Denis Hilton, Joshua Knobe and Yoshi Kashima. Thank you to Victoria University for providing funding through my PhD scholarship.

To my friends who were always there for me even when I didn't return their calls, thank you. A special thanks to Mary who wouldn't let me stray too far for too long and to Rachel, for your positivity and humour.

Finally and most importantly, thank you to my family. To Lauren and Hannah, you have always believed in me, given me unwavering support, and inspired me to aim a little higher. Without your support, I couldn't or wouldn't have wanted to do it, thank you. To Sam, thank you for pondering rabbits and pinecones while I searched for clarity in the Town Belt. To Stephen, all the gratitude in my heart is yours: you gave me a dream and helped make it a reality. To borrow from Newton (1664): If I have seen further it is by standing on your shoulders.

Table of Contents

Acknowledgments	i
Table of Contents	iii
Abstract	v
Chapter One	1
Legal Theories of Causation	3
Psychological Research on Attributions in Chains of Multiple Events	4
Attributional Theories of Causation	6
Positive Actions and Outcomes: Effects of Positive or Negative Valence on Attributions. 13	
Valenced Events and Physical Causes in Chains with Positive or Negative Outcomes.	15
Overview of the Experimental Paradigm.	15
Overview of the Research Hypotheses	17
Chapter Two: Study 1 - Human Agency, Motive and Outcome Information	23
Chapter Three: Explaining the Effects: Choice of Measures or Outcome Valence?	33
Study 2: Extending the Measures of Praise/Blame and Responsibility	33
Study 3: Adding Measures of Punitiveness and Probability	42
General Discussion	53
Chapter Four: Explaining the Effects: Abnormality or the Facilitative Framing of Events?	57
Study 4: Does Abnormality Explain Attributions?	60
Study 5: The Effect of the Facilitative Nature of the Event.	69
General Discussion	76
Chapter Five: The Role of Intentionality and Foresight on Attributions	79
A New Attributional Theory: Sloman, Fernbach and Ewing's (2012) Causal Model	84
Study 6: Testing the Role of Intentionality and Foresight	86
Chapter 6: General Discussion	107
Review of Findings	107
Two Patterns of Judgments: A Mismatch Effect and a Motive Effect	110
The Relations between Constructs of Cause, Responsibility and Blame	111
Implications for Psychological and Legal Theories of Causation	112
Attributional Theories of Perceived Causation.	113
Broader Methodological Issues and Future Directions	117
Conclusion	121

References.....	123
Appendix A: Scenarios	132
Studies 1-3	132
Study 4.....	136
Study 5.....	146
Study 6.....	151
Appendix B: Study 1 – Descriptive Statistics	163
Appendix C: Study 2 – Descriptive Statistics	164
Appendix D: Study 2 –Relations between Judgments of Cause, Blame, and Responsibility	166
Appendix E: Study 3 – Descriptive Statistics.....	168
Appendix F: Study 4 - Pilot Studies.....	170
Pilot Study 1: The influence of two different types of motivation on attributions.....	170
Pilot Study 2: The influence of abnormality, and two types of physical events on attributions.	171
Appendix G: Study 4 – Descriptive Statistics	175
Appendix H: Study 4 - Tests of Legal Theories of Causation.	177
Appendix I: Study 5 – Descriptive Statistics	178
Appendix J: Study 6 – Descriptive Statistics	179

Abstract

Research on attributions about several events in causal chains has focused on chains ending in negative outcomes and has not examined positive outcomes and actions (e.g., Hilton, McClure, & Sutton, 2010; Lagnado & Channon, 2008; McClure, Hilton, & Sutton, 2007). On the other hand, research on attributions for positive and negative events has examined judgments about one event in the chain and has not examined effects on other causes in the chain or made comparative judgments about physical causes that produce similar effects to actions (e.g., Alicke, 1992; Alicke, Rose, & Bloom, 2011). This thesis integrates these two lines of research. Six studies examined judgments about two consecutive events (intentional actions and physical events) in chains leading to positive as well as negative outcomes. The intentional action was the same action (e.g. a man started a fire) that differed in motive (positive or negative). The physical event had the same causal effect as the action (e.g., a lightning strike started a fire), or was a physical event (e.g., strong wind) that occurred later in the causal chain.

The results replicate previous findings that when both actions and outcomes are negative, participants rate intentional actions more causal and blameworthy than physical events. However, when the intended outcomes fail to eventuate or positively motivated actions pre-empt positive outcomes, two distinct patterns emerged: A *mismatch effect* that explains the cause of the outcome; and a *motive effect* that explains judgments of culpability (measured by judgments of blame and punitiveness in these studies). Specifically, judgments of cause, responsibility, intentionality and foresight follow the same pattern that reflects the congruence between the valence of the agent's motive and the outcome. In contrast, judgments of culpability follow a different pattern where motive and outcome information have independent effects. Notably, it is the moral intent of actions that primarily determines judgments of culpability. The valence of the outcome plays a secondary role and amplifies ratings.

These results show that the important psychological and legal concepts of intentionality, abnormality, foresight, proximity, and outcome information are core determinants in lay attributions (e.g., Hart & Honoré, 1985; Heider, 1958; Kelley, 1973; Weiner, 1995). But it is valence that plays the critical role in shaping lay reasoning. Several theoretical approaches applied in previous research on causal chains are examined, for example, Alicke's (2000) culpable control model, Tetlock's (2002) social functionalist model, and Spellman's (1997) crediting causality model. Yet none of the theories are able to account for the findings for chains that include positive actions or positive outcomes. The theoretical scope of this thesis was

expanded in Study 6 to include research on the folk concept of intentionality, hindsight, and actor-observer biases (Fischhoff, 1975; Kashima, McKintyre, & Clifford, 1998; Malle & Knobe, 1997; Malle, Knobe, & Nelson, 2007). The results are interpreted in terms of Sloman, Fernbach and Ewings' (2012) causal model of intentionality.

Chapter One

Does knowledge of another's motive in a chain of events influence our judgments of the causal effect of that action and whether the person is held culpable? If so, are these effects moderated by the action or the outcome of the chain? For example, a man enters a forest and deliberately starts a fire. In one case, the man is an arsonist whose goal is to destroy as many homes as possible. In the other case, the man is a Conservation Department officer whose goal is to prevent a wildfire from spreading to the nearby town¹. In cases like this, do we recognise both men's actions as having the same causal effect despite differing goals? Or, do people's good intentions (in this case, the officer's desire to save the town from a wildfire) buffer them from blame if the outcome goes awry? Alternatively, do we judge the arsonist as harshly when the intended outcome doesn't occur (i.e., the homes weren't destroyed)? If a subsequent breeze sprang up and fanned the flames, will people revise their assessment of the man's causal effect? Or, do people consider physical events that occur later in the causal chain (e.g., the breeze) as background conditions when intended outcomes eventuate but as significant causes when unintended consequences occur?

As the concept of causation is fundamental to our understanding of the world around us and for our ascription of moral and legal responsibility, unravelling the complexities of how people make attributions has been the subject of considerable interest in social psychology, law and philosophy. Indeed, important legal concepts such as *mens rea* (e.g., Hart & Honoré, 1985; Moore, 2009) and philosophical ideas regarding causal relations (White, 1990) have been examined and incorporated into attributional theory (e.g., Heider, 1958; Kahneman & Miller, 1986; Kelley, 1967; Malle & Knobe, 1997; Weiner, 1995). Researchers have shown that in order to explain the action of others, people seek answers to the fundamental causal questions: what, who, why and how? (e.g., Abelson & Lalljee, 1988; Hilton, 1990; Kun & Weiner, 1973). People make causal inferences and create mental representations of causal sequences that lead to an outcome (Kahneman & Tversky, 1984; Weiner, 1995). People consider what factors might have caused the outcome, what the alternatives were, who or what was responsible, and whether those actions warrant reward or punishment (Alicke, Davis, & Pezzo, 1994; Alicke, Yurak, &

¹ This vignette is adapted from Hart and Honoré's (1958, p. 71) well-known forest fire scenario and will be used as an illustrative example throughout this thesis. The lack of differentiation between cause (a small fire) and effect (a large fire) is a limitation. However, the vignette has been adapted by several authors in closely related research (e.g., Hilton & McClure, 2012; Lagnado & Channon, 2008; McClure, Hilton & Sutton, 2007) who found no important differences between scenarios therefore supporting its inclusion in these studies.

Vredenburg, 1996; Epstude & Roese, 2008; Hilton, McClure, & Slugoski, 2005; McClure et al., 2007; Tetlock, 2002; Woolfolk, Doris, & Darley, 2006).

These questions, whether raised in a court of law or the workplace, have generated much research on the cognitive processes and other factors that people utilise to make judgments of cause and blame. The findings support the special status that psychological and legal theorists ascribe to human agency and the conditions determining whether an event is identified as the primary cause of an outcome (e.g., Hart & Honoré, 1985; Heider, 1958, Kelley, 1967; Malle & Knobe 1997). Specifically, several factors consistently emerge as core determinants in attributional decisions. These include actions performed with intent (Heider, 1958; Malle & Knobe, 1997), control (N'Gbala & Branscombe, 1995), foresight (Fischhoff, 1975; Schkade & Kilbourne, 1991), awareness (Malle & Knobe, 1997), and consequential knowledge (Baron & Hershey, 1988; Enzle & Hawkins, 1992). Other factors include the event's proximity to the outcome (Alicke, 1992; Segura & McCloy, 2003; Wells, Taylor, & Turtle, 1987), and the severity of the outcome (McClure, Lalljee, & Jaspars, 1991; Tetlock, Self, & Singh, 2010). Recently, studies examining causality in more complex narrative sequences have shown that patterns of attributions are further differentiated depending on the causal structure and content of the vignette (e.g., Alicke, 1992; Alicke, Buckingham, Zell, & Davis, 2008; Alicke et al., 2011; Catellani, Alberici, & Milesi, 2004; Hitchcock & Knobe, 2009; Tostain & Lebreuilly, 2008), the type of attribution being sought (e.g., Cushman, 2008; Tetlock et al., 2007), and the cognitive processes invoked (e.g., Cushman, 2008; Mandel, 2003).

These studies demonstrate that attributing causality is complicated in practice and that a myriad of events could be cited as causes of a particular outcome. The inconsistent findings suggest, however, that these studies are constrained by two major limitations. First, many of these studies used simple causal structures that present covariation information or offer 'either-or' choices without narrative context, so their findings may not generalise to more complex scenarios in the 'real' world. Second, these studies elicited judgments for a single target event, so they tell us little about how the temporal order of the core determinants (outcome, abnormality, foresight, etc.) influence attributions. These limitations raise questions about the conditions necessary for a single event to be seen as the primary cause of the outcome. Specifically, how do the various inferences of intentionality, foresight, control, and proximity relate to one another? Are the relative weightings and the relations between factors applied consistently, or are they contextually dependent?

In the remainder of this chapter, I discuss perceived causation in legal theory and review psychological research that has tested three legal hypotheses in sequences of multiple events. I then describe the two dominant approaches to attributional research in social psychology before reviewing three theories in psychology that have been applied to attributions in causal chains. Finally, as the role of valence has not been examined in studies with multiple events, I review research that examined the influence of action and outcome valence on attributions for single target events.

Legal Theories of Causation

Legal theory and in particular the doctrines of proximate and intervening causation, provide a useful framework for researching lay attributions for multiple events in chains comprising voluntary actions and physical events. Due to the importance of decisions of causal responsibility and culpability in courts of law, tracing the causal path from the outcome to the defendant is particularly important in legal reasoning. When selecting the primary cause of an outcome from a sequence of events, Hart and Honoré (1985) proposed that there are only two conditions ‘eligible to be identified’ as the primary cause: free, informed, voluntary human actions and abnormal natural events; any remaining conditions are considered background or ‘mere’ conditions (p. 72). Hart and Honoré also claim that causality cannot be traced back through time indefinitely. Instead, the causal relation between an event and the outcome either diminishes by degrees over the number of events, or is suddenly ended by an intervening event². The intervening event however, can only ‘break’ the causal contribution of an otherwise primary cause if it also is one of the two eligible conditions.

Accordingly, Hart and Honoré (1985) propose first, that people prefer voluntary causes over physical events as causal explanations; and second, that in a sequence of events people attribute the outcome to the most recent voluntary occurrence. For example, when the chain comprises of a voluntary action (e.g. a man deliberately lighting a fire) followed by a later physical event (e.g., a breeze), people see the voluntary action as the cause. In contrast, if the voluntary action falls later in the chain of events (e.g., a man deliberately fanning the flames) and takes advantage of a prior physical event (e.g., sun focused by a piece of broken glass set fire to a shrub), people see the later action as the primary cause and relegate the prior event to the background. Third, in chains of two comparable events (e.g., two voluntary actions, or two

² Referred to as the doctrines of proximate and intervening causation in legal theory (Moore, 2009, pp. 118-129).

physical events), the most recent event is considered the cause. For example, if one man had deliberately started the fire, but a second man who was unaware of how the flames were caused, deliberately fanned the flames, the latter, independent action is seen as the primary cause.

Hart and Honoré's (1985) criterion of abnormality however, implies that in chains of two physical events, abnormality is more important than proximity. For example, if lightning started the fire and a breeze fanned the flames, the lightning strike is seen as the cause, being the more abnormal event. But, if unseasonably strong winds fanned the flames of a fire started by either lightning or focused sunlight, the wind is considered the cause, being the more recent abnormal event. Finally, the theory proposes that an abnormal physical event is judged the intervening cause if that event is causally independent of the voluntary action. For example, if unseasonably strong wind gusts fanned the flames of a deliberately lit fire, Hart and Honoré claim that the wind, being the later abnormal condition, would be seen as the cause of the fire. If, however, the man knew of the predicted abnormally strong winds, his action would be treated as a 'contrived coincidence' and deemed the cause.

With regard to the relationship between cause, blame, and responsibility, Hart and Honoré (1985) proposed that actions that are seen as the cause of outcomes are 'blamed' only when they result in harm or people infer liability. In contrast, responsibility does not require a causal connection to an effect but is ascribed when a person is held liable for the outcome. For example, a toddler playing with matches might be the cause of a fire, but the caregiver would be held responsible for the child's actions. To differentiate how intent, foresight, and proximity relate to outcomes, Hart and Honoré distinguished three aspects of blame: Direct blame, where actions are intentional and have foreseen harmful effects; indirect blame, where unintentional or negligent actions result in unforeseen harmful consequences; and moral blame, where there is intent without a harmful consequence, such as inchoate offences where intentional actions result in unforeseen outcomes.

Psychological Research on Attributions in Chains of Multiple Events

Several psychological studies have drawn on Hart and Honoré's (1985) legal theory of causation as a conceptual framework to examine how people determine culpability and select an event as the primary cause while relegating others to background conditions in chains of events (e.g., Hilton & McClure, 2013; Hilton et al., 2010; McClure et al., 2007). These studies describe a sequence of independent events that lead to a harmful outcome. For example, McClure et al. (2007) adapted Hart and Honoré's (1985) fire scenario and two other scenarios to study

opportunity chains (Hilton et al., 2005), where the occurrence of the first event (e.g., a shrub catches fire) enables a second event (e.g., the flames being fanned) which leads to a harmful outcome (e.g., a forest fire). Participants rated the causality of events in the first (distal) and second (proximal) position in chains with voluntary actions (e.g., a man deliberately set fire to the shrub) or physical events (e.g. sunlight focused by a piece of glass set fire to the shrub).

Research has supported several of Hart and Honoré's (1985) claims and shows that the temporal position (distal or proximal) and type of events (voluntary or physical) influence attributions (Hilton et al., 2010; Lagnado & Channon, 2008; McClure et al., 2007). As the theory predicts, first, people rate intentional actions (e.g., a youth setting fire to a shrub) higher than physical events (e.g., the sun setting fire to a shrub) as causes of harmful outcomes. Second, when the chain comprises voluntary actions and physical events, irrespective of their position in the chain, people see the voluntary actions as the primary cause and discount the physical events (Hilton et al., 2010; Lagnado & Channon, 2008; McClure et al., 2007).

However, the findings lend only qualified support to the claim that when a sequence comprises two comparable events, people see the most recent abnormal event as the cause. Lagnado and Channon (2008) found that people see the latter event as more causal when the chain consisted of two physical events, but not when the chain consisted of two voluntary actions, where both actions were rated equally causal. McClure et al. (2007) found that in chains of two voluntary actions or two physical events, people rated both events as equally good explanations of the outcome.

The findings are also inconsistent when a distinction is drawn between intentional and unintentional actions. Lagnado and Channon (2008) found that people rated cause and blame higher for intentional actions than unintentional actions and physical events which did not differ. In contrast, Hilton et al. (2010) found no difference between deliberate and non-deliberate human actions; both were rated more causal than physical events. However, ambiguous distinctions between the types of event may have confounded responses. For example, the researchers' description of a proximal physical event (e.g., a breeze) may not have been considered sufficiently abnormal to 'break' the causal effect of the preceding physical event (e.g., the sun setting fire to a shrub). Furthermore, an action defined as unintentional (e.g., a man spraying water on a road in zero degrees without thinking) may have been seen as negligent and the consequences foreseeable rather than unintentional *per se*, so participants' attributions were more closely aligned with those for intentional actions.

The evidence supports Hart and Honoré's (1985) claim that voluntary actions are 'special', yet the findings also suggest that lay attributions do not always follow legal principles when applying information about intentionality and the temporal order of events. Instead, people may be more responsive to other implicit factors rather than those explicitly stated in the vignette. For example, people may be attending to implications of uniqueness and the sufficiency of the event to cause the outcome (Spellman, 1997), their own moral (Knobe, 2010) or evaluative reactions (Alicke, 2000) to the events, or the extent that the actions are informative (Uttich & Lombrozo, 2010) or violate social norms (Tetlock, 2002).

Several studies have examined the conceptual distinctions people draw between attributions of cause, blame, punitiveness and wrongness (e.g., Alicke et al., 2011; Cushman, 2008; Tetlock et al., 2007). Overall, the findings show that people distinguish between these constructs as prescribed by legal theorists but reveal several anomalies. For example, the ascription of blame presupposes that harm has been risked or realised (e.g., the action was intentionally harmful, malevolent, or negligent), whereas responsibility, being a more semantically neutral term, reflects people's perceptions of the agent's accountability for either positive or negative outcomes. Thus, valence of the outcome should have a diverging influence on judgments of blame but not responsibility (Alicke, 1992; Knobe 2010; Weiner, 1995). Yet research has shown asymmetries for outcome valence for blame and responsibility, suggesting that other factors are influencing lay interpretations (Alicke et al., 2008; Alicke et al., 1996; Enzle & Hawkins, 1992; Malle, 2006a).

Attributional Theories of Causation

In psychology, two theoretical approaches have dominated how people conceptualise and explain their own and others' action: probabilistic theories such as the covariational approach stemming from Kelley's (1967) model of attribution (e.g., Cheng & Novick, 1991; Spellman, 1997), and the functionalist approach based on Heider's (1958) theory of causation (e.g., Alicke, 2000; Tetlock, 2002). Probabilistic approaches to causal attribution proposes that people reason as 'intuitive statisticians' and any causal effects are explicable in terms of probability (e.g., Cheng & Novick, 1991; Kelley, 1967). In contrast, the functionalist approach claims that people explain actions in terms of reasons for acting, and focus on the unique properties of intentional action. Both approaches are similar in their treatment of physical events and unintentional action; unintentional actions are comparable to any other event and are explained in terms of cause-

effect, but they take very different views of intentional action. Researchers have applied both approaches to research on causal chains.

The covariational approach to lay attributions.

The covariational approach claims firstly, that people see intentional action as events that they explain in terms of the person, the stimulus, or the occasion. Secondly, people analyze covariation information and attribute cause to ‘that condition which is present when the effect is present and which is absent when the effect is absent’ (Kelley, 1967, p. 194). Thus, the causal effect of each event is calculated as the perceived probability of the outcome given the presence (or sufficiency) of the cause [$P(O|n+e)$] less the absence (or necessity) of the cause [$P(O|n \sim e)$].

Cheng and Novick’s (1991) probabilistic contrast model claims that people ascribe causality to the event that most increases the probability of the outcome. Spellman (1997) extended it to sequences of events and proposed that causal selection is based not on discrete analyses of individual events, but a comparative process that is conditional on what has gone before and what transpired. Thus, similar to Hilton’s (1988, p. 52) claim that people conduct a ‘naïve form of multiple regression’ to evaluate potential causes, Spellman proposes that causal selection is based on estimates of conditionalised sufficiency, where cause is ascribed to the event that increases the probability of the outcome the most less the probability of the outcome prior to the target event’s occurrence: mathematically shown as $[P(O|n+e) - P(O|n)]$.

One implication of the probabilistic models is that the type of event should not necessarily influence causal attributions. In Spellman’s model (1997), the event that raises the probability of the outcome the most is considered the cause of the outcome, regardless of whether it is a voluntary action or a physical event. Thus, any causal effects should be explicable in terms of differences in probabilistic contribution to the outcome. Returning to the forest fire scenario, for example: if a lightning strike and a man start the same sized fire; both events equally increase the probability of the outcome (a shrub catching fire) and causal attributions between these events should not differ. Therefore, if intentional actions are seen as more causal than physical events as suggested by legal theory, it is because they are stronger covariates of the outcome. Indeed, Spellman (1997) drew on Heider’s (1958) property of equifinality and notion of ‘effective personal force’ (p. 82) as a probabilistic explanation for people’s preference for voluntary actions³. Heider argues that intentional actions are judged more causal than unintentional actions

³ Heider’s (1958) theory of causality focused on the intentionality of actions. According to Heider, intentional actions (which he referred to as personal causation) are goal-directed and characterised by the property of equifinality where the same effect is achieved despite a change in the means. In contrast, unintentional actions and physical events (which he referred to as impersonal causality) are not goal-

or physical events because they imply the agent exercised a degree of control and co-ordination of conditions to cause an invariant outcome.

Spellman (1997) found support for the crediting causality model in sequences that manipulated probabilistic success with coin tosses or simple games. However, support has been mixed when researchers tested Spellman's model in more naturalistic situations (e.g., Hilton et al., 2010; Lagnado & Channon, 2008; Mandel, 2003; McClure et al., 2007). The lack of consistency however, might reflect differences in the type of chains described in the scenarios chosen (see Hilton, 2005). Hilton et al. (2010) compared how well probability principles explained causal attributions in unfolding chains where both events are necessary conditions that lead foreseeably to the outcome (e.g., an accident). For example, the proximal physical event (e.g., ice on the road), was dependent on the distal event's occurrence (e.g., someone spraying water on the road) in a zero temperature. The study tested the principles of conditionalised and unconditionalised sufficiency (the outcome always occurs when the target event is present) and covariation (the outcome always occurs when the target event is present and never occurs when the target event is absent) as predictors of causal attributions⁴. The results showed that participants used different sufficiency principles depending on the type of target event, and secondly, that unconditionalised sufficiency rules best predicted causal selection in sequences of multiple events; whereas conditionalised sufficiency rules (e.g., Spellman, 1997) added predictive value when the distal events were voluntary actions but not physical events.

In contrast to the unfolding chains used by Hilton et al. (2010), McClure et al. (2007) and Lagnado and Channon (2008) examined probability reasoning in opportunity chains. In these chains, the distal event (e.g., a youth, or sun setting fire to a shrub) does not cause, but enables the later, independent event (e.g., a breeze, or a man fanned the flames) to produce a negative outcome. Both studies found that people prefer intentional actions to physical events as the primary cause, but neither study supported Spellman's (1997) proposal that conditionalised statistical principles mediate the relationships between the differing types of events and the outcome. In these chains, voluntary actions did not add explanatory value by virtue of their independent effect on probability of the outcome.

directed and thus are characterised by the property of multifinality, where a change in means leads to a change in effect.

⁴ Conditionalised analyses applies the principles of covariation and sufficiency to causal chains rather than unconditionalised analyses which focuses on single target events (Hilton et al., 2010, p. 388). Conditionalised covariation is the covariation between the event and the outcome less the covariation of the previous event in the chain and that outcome. Conditionalised sufficiency is the sufficiency of the target event less the sufficiency of the previous event (e.g., Spellman, 1997).

The finding that sufficiency rules explain causal attributions in unfolding chains but not opportunity chains, suggests that causal reasoning is influenced by the structure of the causal chain. Thus, the presence or absence of constraints between two successive events appears to influence the causal weightings people apply to the central features of voluntariness, abnormality, intent, control, and foresight. It is as if people exercise cognitive economy by ignoring or discounting the probabilistic contribution of physical events in favour of preceding voluntary actions when the sequence proceeds in a relatively systematic order as in unfolding chains (Heider, 1958). In contrast, when the sequence is constrained by the relations between events (as in opportunity chains), people appear to take a more holistic approach in considering the necessity and sufficiency of each condition and the causal weightings they attach to each factor. In sum, these findings support the legal premises that voluntary actions are ‘special’, and that causal attributions are not independent, but are influenced by what has gone before and after each event. However, the findings also qualify probabilistic theory by suggesting that people’s attributions are not primarily determined by statistical principles.

The functionalist approach to lay attribution.

Several functionalist approaches have been applied to attributions about causal chains (Hilton & McClure, 2010, 2013; Lagnado & Channon, 2008; McClure et al., 2007; Tetlock, 2002; Tetlock et al., 2007). However, only two theories have emerged as offering succinct explanations of the findings on attributions in causal chains: Tetlock’s (2002) social functionalist model and Alicke’s (2000) culpable control model, to which I now turn.

The Social Functionalist Model.

Tetlock’s (2002) social functionalist model claims that attributions reflect societal controls and that people act as ‘intuitive prosecutors’ to defend the normative order. According to this model, the attributional process is initiated when people perceive a shared social or moral norm has been violated and they consider it worthy of social sanction through deterrence, prevention, or retribution.

Recently, Tetlock et al. (2007) proposed the ‘fair-but-biased-yet-correctible’ (FBC) model as an extension of the social functionalist model to incorporate two groups of observer factors that influence attributions. First, the model posits the ‘prosecutorial mindset’ is activated when the observer perceives a norm has been intentionally violated. The ‘prosecutorial mindset’ is a cluster of cognitive and emotional factors that inform the observer’s sense of moral outrage, the severity of punishment, along with character inferences that shape their assessment of foresight, intentionality and duress.

Second, the model proposes that people's responses to norm violations are moderated by three assumptions: judgments are inherently biased, people prescribe to the concept of fair play and normative justice, and people undertake self-monitoring and correction to mitigate overt harshness and biases. Thus, upon activation of the prosecutorial mindset, people weigh the severity of the outcome, together with any excuses, justifications, or extenuating circumstances against the actor's perceived control over the outcome. However, it is the observer's predilection toward defending social norms that leads them to justify their taking factors into account that are beyond the violator's control. For example, if the forest fire destroyed dozens of homes following a sequence of physical events, the model predicts that people would evaluate those events' causal effects for their utility for future preventive intervention. In contrast, if the sequence contained voluntary actions (e.g., a man lighting a fire) and physical events (e.g., wind), the voluntary actions are seen as the cause because they offer greater social controllability through sanctions and retribution.

Finally, the model offers diverging predictions in regards to motive. When the motive is harmful (e.g., the arsonist whose goal was to create a wildfire), the model proposes that the prosecutorial mindset is activated and retribution is amplified to a level where the norm violations have been appropriately redressed. In contrast, with positive motives (e.g., the Conservation officer whose goal was to prevent the wildfire from spreading), the fairness and self-correction postulates suggest that positive intent is recognised as mitigating evidence, so judgments of retribution are suppressed in favour of future-oriented interventions. Finally, the model suggests that negative outcomes amplify attributions whereas positive outcomes attenuate attributions.

Tetlock et al. (2007) found support for the premise that norm violations activate a prosecutorial mindset which is moderated by the observer's ideology. In a series of studies, the researchers manipulated societal threat, ideology (liberal, moderate, or conservative), excludability, defendant status, outcome severity, and adequacy of punishment. Participants read a story describing a city that was either very law abiding, or had very high crime but low conviction rates, followed by a scenario describing an aggravated robbery which resulted in the victim's death or in minor injuries. Participants made initial attributions before reading an explanation of the defendants' action ranging from extremely extenuating (e.g., the defendant suffered serious neurological problems), to extremely exacerbating (e.g., defendant planned to spend the money on drugs and prostitutes). After reading the explanation, participants could revise their attributions. The results show that conservatives were more punitive and attributed

more responsibility for outcomes when the conditions were ambiguous (e.g. with moderate extenuating or exacerbating circumstances) than liberals. However, conservatives and liberals alike, attributed greater blame to defendants in higher status roles who committed negligent acts with serious consequences, but only conservatives did so for low-status defendants. In a similar study, Tetlock et al. (2010) examined the effects of extenuating information and outcome severity in two cultures; American and Singaporean. The results showed cultural variations in how people adhered to social functionalism when attributing retribution for norm violations, supporting the claim that societal norms play a key role in attributions.

Both studies by Tetlock and colleagues successfully applied the social functionalist framework to attributions for a single target event. However, the findings are inconsistent when Tetlock's (2002) model has been applied to chains comprising voluntary and physical events (e.g., Hilton & McClure, 2013; McClure et al., 2007). McClure et al. (2007) found that irrespective of temporal order of the different types of events, judgments of social controllability mediated causal judgments. Hilton and McClure extended this research by distinguishing between the epistemic state of the voluntary action (the agent either knowingly or unknowingly took advantage of an opportunity to cause harm) and physical events. In contrast to McClure et al.'s findings, Hilton and McClure found social controllability did not explain the causal preference of voluntary actions over physical events.

The Culpable Control Model.

The second key psychological model applied to causal chains is Alicke's (2000) culpable control model. The model's core premise is that people's evaluative reactions to others' action, the events that surround them and the consequences influence judgments of blame, which in turn shape judgments of cause. Thus, Alicke's model opposes the view that causal judgments determine blame attributions (e.g., Weiner, 1995), and instead proposes that people's automatic evaluative responses lead to an 'initial blame hypothesis' (Alicke et al., 2010, p. 9). If the source of the evaluation is judged sufficiently negative, a 'blame validation mode' is activated, leading people to assess the actor's personal control in a way that supports their initial blame hypothesis. The model thus differentiates between judgments of blame and cause, for which it makes diverging predictions depending on the observer's perception of the agent, the events, and the outcome.

Therefore, Alicke (2000) claims that two simultaneous processes are involved in attributions: first, people take a deliberative approach to ascertain the agent's actual and potential control over the events. This includes assessing three aspects of control: the agent's behavioural

control (the extent the action was freely performed and intentional), causal control (mechanical link to the outcome) and outcome control (causing the outcome in the foreseen and desired manner). Second, people have spontaneous affective reactions to the agent, the event, the outcome, and to the 'peripheral features of the event' (e.g., gender, ethnicity, character, outcome severity). The model also proposes that when there is contextual ambiguity and people are unsure of the agent's intentional involvement when harm eventuates, a compensatory response is provoked and other causal explanations are evaluated. In cases such as this, Alicke claims that the observers' negative affective reactions to the outcome and heightened desire to blame, results in the agent's causal control being discounted in favour of alternative events.

Applying Alicke's model to the fire scenario, for example: the arsonist's negative motive provides a strong basis for people to assess the agent as blameworthy; the valence of the outcome then provides evidence that either justifies or attenuates that desire to blame. If the forest fire led to the destruction of dozens of homes, the negative outcome is seen as an aggravating circumstance that further elevates the arsonist's perceived control (behavioural, causal and outcome control) beyond the level if the harm had not occurred, or indeed a positive outcome had eventuated. In contrast, if the agent acted with a positive motive (e.g., the Conservation officer) but a harmful outcome nevertheless occurred, the model suggests a negative response to the outcome would provoke a compensatory response because the officer's actions are beyond reproach. As such, the baseline level of perceived behavioural and causal control attributed to the agent remains unchanged, but the causal effect of the proximal physical event (e.g., the wind) will be amplified to justify the desired blame judgment. If, however, people saw the officer as having acted negligently when the outcome went awry; their negative response to the harmful consequences coupled with a basis for negative evaluations of the action would lead them to amplify assessments of the officer's control.

The culpable control model is supported by a number of studies (Alicke, 1992; Alicke et al., 2008; Alicke et al., 1994; Alicke et al., 2011; Lagnado & Channon, 2008; Mazzocco, Alicke, & Davis, 2004). Recently, studies have shown people's evaluative reactions rather than moral or norm violations best explain attributions in sequences that involve negative action. For example, Alicke et al. (2011, Expt. 2) presented participants with a scenario that differed in the normative action of the group (all of the students either cheated on the test or did not cheat), and the target's action (Granger cheated or did not cheat). When Granger cheated on the test, adhering to group norms made no difference to either cause or blame judgments; his blameworthiness and causal influence was amplified regardless of the other students' action. In contrast, adhering to group

norms influenced ratings when Granger refrained from cheating; when everyone else cheated Granger's blameworthiness and causal influence was attenuated, but not when they all upheld the normative standard. These findings imply that while norms are important, it is not because they provide an effective form of social control (e.g. Tetlock et al., 2007) but because they provide a evaluative gauge for people to compare the actual action against.

The core difference between Alicke's (2000) and Tetlock's (2002) theories are their view of 'what' drives people's judgments of events. Tetlock's model maintains that a norm (social or moral) violation is the key influence in causal judgments. In contrast, Alicke's model argues that norm violations alone are not sufficient to account for causal attributions, instead proposing that people's evaluative reactions to the agent, outcome and events (social and moral norms being one aspect of these) have the more pervasive influence on judgments.

Positive Actions and Outcomes: Effects of Positive or Negative Valence on Attributions.

The research examining judgments about two or more causes in causal chains has followed the legal framework and focused on negative actions and outcomes, yet the authors have stated conclusions about causal chains that logically apply to all voluntary actions and outcomes, not just negative events (Hilton et al., 2010; Lagnado & Channon, 2008; Mandel, 2003; McClure et al., 2007). However, a large body of research showing that negative events have greater impact than positive events across a number of domains suggests this generalisation to positive events might be unwarranted (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Ybarra, 2002).

Effects of the valence of the outcome on attributions.

Research examining judgments about single events has shown that the positive or negative valence of the outcome triggers very different judgments (Baron & Hershey, 1988; Bohnet, Bless, Schwarz, & Strack, 1988; McClure et al., 1991; Robbennolt, 2000; Tostain & Lebreuil, 2008). When asked to evaluate events leading to positive or negative outcomes, people amplify ratings of intentionality, cause, blame, and responsibility more for negative than positive outcomes (Alicke et al., 1996; Alicke et al., 2008; Kahneman & Tversky, 1984; Malle & Knobe, 1997; Mazzocco et al., 2004). Linking this effect to counterfactual processing, Alicke et al. (2008) found that people ascribe more blame when the outcome could easily have been avoided and the action was within the actor's control. In addition, the more similar an actual outcome was to an imagined favourable outcome, the lower the blame (Shaw, Wild, & Colquitt, 2003). This outcome-valence asymmetry has been replicated in studies examining the neural and behavioural mechanisms in moral judgments (e.g., Young, Cushman, Hauser, & Saxe, 2007) and

the role of emotion in human actions (e.g., Forgas, 2008; Zhu & Thagard, 2002), suggesting the effect is not limited to a cognitive bias.

Alicke et al. (2011, Expt. 1) extended this research by examining attributions in chains with positive and negative actions and outcomes. Expanding Hitchcock and Knobe's (2009) physician scenario, Alicke et al. described a sequence of events where the physician's action was either normative (e.g., he follows hospital policy forbidding the use of this drug and refuses to administer a drug) or counternormative (e.g., despite hospital policy forbidding the use of this drug, he administers a drug). The patient then experienced either a positive outcome (e.g., the patient recovers) or negative outcome (e.g., the patient dies). Alicke et al. found that participants saw actions that violated procedural norms most negatively when the outcome was negative, but when the action followed procedural norms their judgments of cause did not differ depending on the valence of the outcome. These findings replicate earlier studies demonstrating asymmetries in judgments for positive or negative events and suggest that attributions follow the predictions of legal theory (e.g., Hart and Honoré, 1985) when the sequence leads to negative outcomes but not positive outcomes. Instead, these findings show that people's evaluative reactions to the valence of the action and the outcome interact to have a differing effect on attributions. When things turn out well, adhering to normative standards influences judgments of morality but has little effect on judgments of cause. In contrast, when things go wrong and harm ensues, following normative standards shapes judgments of both culpability and causation.

Finally, the strength of the asymmetry in attributions due to outcome valence extends beyond direct effects. For example, Knobe (2003) showed that the valence of foreseen secondary outcomes, or side-effects influenced judgments of intentionality despite the primary outcome being attained. Participants read that the chairman of a company was told about a new program that will help increase profits, but will also *harm* (or *help*) the environment. The chairman responds 'I don't care at all about *harming* (or *helping*) the environment. I just want to make as much profit as I can. Let's start the new program.' The new program begins and sure enough the environment was *harmed* (*helped*). When asked if the chairman *intentionally harmed* (or *helped*) the environment, most participants judged 'yes' the chairman acted intentionally when the side-effect was negative, but not when the side-effect was positive. This asymmetry in intentionality judgments known as the 'side-effect effect' has been replicated several times and extended to different methods (e.g., Cushman, Knobe, & Sinnott-Armstrong, 2008; Guglielmo & Malle, 2010b), and scenarios (e.g., Machery, 2008; Mallon, 2008; Uttich & Lombrozo, 2010).

The effects of the valence of the motive on attributions.

Research has shown that the valence of the motive behind an action affects causal judgments, and particularly the assignment of blame. For example, people ascribed more blame and cause when actions were negative, or actors were malevolent, negligent, or reckless (Alicke, 1992; Alicke et al., 2008; Brockner et al., 2007; Cole Wright & Bengson, 2009; Enzle & Hawkins, 1992). Research on positive and negative actions has also shown that differences in severity, controllability, prior knowledge, justifiability, and degree of intent influence people's judgments of blame and responsibility (Alicke, 1994; Alicke et al., 2008; Enzle & Hawkins, 1992; Fincham & Jaspars, 1979; McClure et al., 1991).

Valenced Events and Physical Causes in Chains with Positive or Negative Outcomes.

Two themes in the literature are fundamental to this thesis: First, research examining judgments about two or more events in causal chains has focused on negative actions and outcomes and has not examined chains with positive actions and outcomes (e.g., Lagnado & Channon, 2008; Hilton et al., 2010; McClure et al., 2007). Second, research on attributions for positive and negative events has examined judgments about one event in the chain and has not examined effects on other causes in the chain or compared with physical events that produce similar affects to actions (e.g., Alicke et al., 2011; Cushman, 2008). These two gaps provide the focus of this thesis and are elaborated on in the next section.

This thesis examines attributions about intentional actions with positive or negative motives and physical causes in chains with positive and negative outcomes. The central aim is to investigate whether judgments of cause and blame can be best understood in terms of the core determinants (i.e., outcome, abnormality, intentionality, foresight, etc.), or whether the positive or negative valence of the agent's motive or the outcome has the greater effect on attributions. In answering this question, this thesis aims to provide a clearer understanding of the processes that people employ when making attributional decisions. This question is not only theoretically interesting but has applied relevance in clarifying the moral, ethical and fiscal ramifications of judgments about a person's action, whether in the court of law or in the workplace.

Overview of the Experimental Paradigm.

The present studies integrate two lines of research: first, they extend research on attributions about causal chains with negative actions and outcomes to include chains with positive actions and outcomes. Second, they extend research on judgments about a single event in the chain to judgments about two consecutive events in the causal chain. This goal is achieved by

examining judgments about two consecutive events (voluntary actions and physical events) in chains leading to positive and negative outcomes. In addition, these studies include judgments about physical events that are substituted for actions and have the same effects as the actions, or are physical events that occur later in the causal chain. Including physical events provides a baseline to compare judgments about human actions in causal chains.

Using the scenario paradigm, these studies examined how participants' judgments about two consecutive events were influenced by the differences in the first event in the chain and the outcome. The first (distal) event in the chain described either an intentional action that differed in motive (positive or negative) or a physical event⁵. The second (proximal) event in the chain was a physical event which was held constant across conditions. The outcome following the two events was positive or negative. For example, with the fire scenario, the distal event was an action (e.g. a man started a fire) with either a positive motive (e.g., a Conservation officer whose goal was to back-burn an area of forest to prevent a larger fire from spreading), or a negative motive (e.g., an arsonist whose goal was to ensure a larger fire spreads), or a physical event (e.g., a lightning strike setting fire to an area of forest). The proximal cause was a physical event,⁶ and the outcome was either positive (saving dozens of homes) or negative (destroying dozens of homes). Participants made judgments about both the distal and proximal target events in the sequence.

These studies replicate earlier research using causal chains described as opportunity chains (McClure & Hilton, 2007; Lagnado & Channon, 2008) rather than unfolding chains (Hilton et al., 2010). One reason why opportunity chains are of interest is that they provide a framework to test psychological and legal predictions (e.g., Hart & Honoré, 1985; Heider, 1958; Moore, 2009; Spellman, 1997) about the conditions necessary for one event to be judged the primary cause of the outcome. Hilton et al. (2005) define an opportunity chain as a consecutive sequence of independent events where the first event (e.g. the shrub catching fire) 'enables' or provides the opportunity for the second event (e.g., the wind) to produce the outcome. Here, both events are equally necessary for the outcome, but the distal event does not necessitate, and therefore does not constrain the occurrence of the proximal event. This allows for a systematic examination of the relations between events, and how the temporal order of the core determinants (voluntariness, abnormality, proximity, foresight, etc.) influence attributions.

⁵ To enhance control across the conditions and eliminate potential confounds in scenario content, the positive and negative motives motivated the same action for all conditions

⁶ The proximal cause was a physical event that was the same for all conditions in each scenario (e.g., strong wind).

The role of intentional actions and motives in attributional judgments has been extensively examined, yet often a precise definition is absent. For this thesis, the operational definitions are based on those proposed by Heider (1958) and McClure (1984). Both authors suggest that an intention focuses on the action the person was trying to do and applies to an action's goal or what they were trying to accomplish, whereas a motive refers to why the person was trying to perform the action and the reasons behind the intentional action (Heider, 1958; McClure, 1984).

Overview of the Research Hypotheses

These studies were designed to test several predictions based on the relevant theories and research. First, I predicted that participants would rate distal events more causal and blameworthy when they comprised intentional actions rather than physical events, as proposed by Hart and Honoré (1985), Alicke (2000), and Tetlock (2002). Second, I predicted that participants would rate negative actions more causal and blameworthy than positive actions, based on the findings that people amplify ratings for negatively valenced events (Alicke, 1992; Alicke et al., 2011; Enzle & Hawkins, 1992).

Third, I predicted that the valence (positive or negative) of the outcome would also influence judgments of cause and blame for both events in the chain, and specifically, that participants would rate the distal and proximal events higher with negative outcomes than positive outcomes. This prediction reflects research showing that people's attributions are higher for negative than positive outcomes (Bohner et al., 1988; Kahneman & Tversky, 1984). This prediction also tests whether Alicke's (2000) and Tetlock's (2002) models that focus on negative events, generalise to attributions for positive events.

These studies were also designed to test the novel predictions derived from Alicke's (2000) and Tetlock's (2002) models. Tetlock's model claims that causal judgments reflect the social value of deterring the future performance of negative actions with sanctions and of rewarding positive actions. Harmful consequences amplify the causal ratings of both positive and negative actions in proportion with the increasing level of necessary retribution. Thus, Tetlock's model predicts firstly, that people would rate negative actions as more causal than positive actions and secondly, that people would rate both positive and negative actions higher when the outcome is negative rather than positive. In contrast, Alicke's model claims that people's response to the agent's motive and the outcome provides two sources of interdependent information. Alicke's model predicts an interaction between the motive and the outcome, where people rate negative

actions as more causal when the outcome is negative than positive, but they rate positive actions as no more causal when the outcome is negative than positive.

Several predictions have also been extrapolated from Alicke's (2000) and Tetlock's (2002) models regarding attributions for consecutive events. Tetlock's model takes a utilitarian approach in its emphasis on the controllability of actions, insofar as these actions offer greater opportunity for intervention than physical events. Therefore Tetlock's model predicts that in chains of actions and physical events, regardless of the temporal order of the events or the agent's motive, people will see voluntary actions as more causal than physical events. Applying Tetlock's model to these studies, I predicted that people would rate distal intentional actions as more causal than proximal physical events.

In contrast, the culpable control model (Alicke, 2000) makes diverging predictions for chains comprising voluntary and physical events depending on the agent's motive. The model suggests that a compensatory relationship exists between negative evaluations and assessments of control based on the assumption that at least a modicum of personal control is required before people ascribe blame to the agent. This assumption implies that if people's desire to blame has been provoked by their negative reaction to harm, the agent's motive then determines how they apportion causal responsibility between the events. For example, a negative motive is likely to provide sufficient evidence to validate their control assumptions; thus proximal physical events will be relegated to the background and their causal contribution to the outcome discounted. In contrast, people are likely to see an agent's positive motive as mitigating evidence, thus the agent's perceived control is likely to be discounted whereas the proximal physical event's causality is amplified. Thus, Alicke's model predicts an interaction between the type of distal event (negative action or positive action) and the temporal position of events (distal or proximal): people will see negative actions as more causal than proximal physical events, but positive actions will be seen as no more causal than proximal physical events.

In sum, when extrapolating predictions for these causal chains, the main difference between the social functionalist and culpable control models concerns the effect of outcome valence. Tetlock's (2002) model predicts main effects for outcome valence (positive or negative), the type of distal event (negative actions, positive actions and physical events), and temporal position (distal voluntary actions or proximal physical events). Thus, people will see events as more causal when the outcome is negative rather than positive; negative actions are seen as more causal than positive actions, which in turn are seen as more causal than physical events; and distal voluntary actions are seen as more causal than proximal physical events. In contrast, Alicke's

(2000) model predicts an interaction between the valence of the outcome and the type of distal event (positive action, negative action or physical event). First, people will see negative actions as more causal than positive actions, which in turn are seen as more causal than physical events. Second, people see negative actions as more causal when the outcome is negative rather than positive, whereas they see positive actions and physical events as no more causal when the outcome is negative rather than positive.

Finally, several predictions examined whether people conceptualise the causal chain as a 'means-end schema' and see the events as forming an integrated sequence, or as a series of independent events that explains their causal attributions, and whether these effects are best explained by legal (Hart & Honoré, 1985) or psychological (Heider, 1958) theories. Hart and Honoré (1985) make two predictions regarding opportunity chains: first, in chains of two equivalent types of events (e.g., two physical events), people see the most recent event as the cause of the outcome. Second, in chains including intentional actions and physical events, people see the most recent intentional action as the primary cause unless the proximal physical is abnormal. Thus, I predicted that in the baseline conditions with two physical events, people will see the proximal event as the cause of the outcome. However, in chains including intentional actions and physical events, I predicted that people will see the distal intentional action as the primary cause of the outcome.

In contrast, as noted previously in the discussion of Heider's model (1958), Heider claims that people's preference for intentional actions over physical events as causes is based on the equifinal properties of the actions. Equifinality is characterised by the intentional action being the one necessary and sufficient condition for the outcome to occur. Thus, Heider suggests that people see a sequence of intentional actions and physical events in a 'gestalt' manner or as a single unified series of events, where the causal effect of all other conditions is discounted in favour of intentional causes. In contrast, multifinality involves unintentional actions and physical events where no one condition is necessary for a particular effect. In this case, Heider suggests people see sequences of two successive physical events as a series of separate incidents, and ascribe cause based on each event's independent effect on the outcome. Therefore, I predicted a main effect for the type of distal event, where participants would see chains including intentional actions and physical events as a single integrated sequence of events leading to the outcome. In contrast, in the baseline condition with two physical events, participants would see the sequence as two independent events.

Overview of the Studies.

This thesis details a series of six experimental studies to clarify how each element in a sequence of events influences people's attributions. The thesis examines how well current psychological and legal theories fare in explaining the pattern of findings that emerges over the six studies. In Chapter 2, Study 1 tested the predictions above with scenarios that depict sequences that commence either with actions with positive or negative motives, or with equivalent physical events, and that end with positive or negative outcomes. In addition to testing the generality of Study 1's findings, each of the following studies were designed to test several novel predictions.

In Chapter 3, two studies examined whether the positive or negative valence effects reported in Study 1 were an artefact of the measures used (Study 2), or the scenarios chosen (Study 3). In Study 2, I examined whether the measure of blame used in Study 1 was inappropriate for moral evaluations of positive actions and physical events, by revising the measure of blame and adding a measure of responsibility. In Study 3, I addressed two questions about the structural equivalence of the outcomes in Study 1. I also tested whether the social functionalist model (e.g., Tetlock et al., 2007) or probabilistic approach (e.g., Spellman, 1997) best explains attributions in these causal chains by examining whether causal judgments are mediated by punitiveness or probabilistic contribution.

In Chapter 4, two studies investigated whether the distinct patterns for judgments of cause and blame generalise across differing contexts, or reflect differences in the exemplars used for each category of events. In Study 4, I tested whether the effects of the abnormality of the events, as proposed by legal theorists (Hart & Honoré, 1985; Moore, 2009) could explain the effects of valence on causal attributions. In Study 5, I examined whether the facilitative framing of the events (Mandel & Vartanian, 2009; Spellman, Price, & Logan, 2001), or participants' own biases accounted for the attributional patterns reported thus far.

In Chapter 5 (Study 6), I expanded the theoretical approach in this thesis to investigate how well current theories of intentionality account for the effects of valence on attributions (Malle & Knobe, 1997; Sloman et al., 2012). I examined whether inferences regarding the agent's intent, intentionality, and foresight explain the asymmetries in attributions of cause and blame that reflect the valence of the agent's motive and the outcome. Table 1.1 presents a summary of each study's design.

Table 1.1

Summary of Study Design

Study	Scenario	Design	Measures
Study 1	3 x Fire, train, factory	2 Outcome (positive, negative) x 3 Distal event (positively motivated action, negatively motivated action, physical event)	<ul style="list-style-type: none"> • Cause • Blame • Unitisation of Chain • Voluntariness (manipulation check)
Study 2	3 x Same as Study 1	Same as Study 1	<ul style="list-style-type: none"> • Cause (same as Study 1) • Praise/blame (bipolar blame measure) • Unitisation of Chain (same as Study 1) • Responsible
Study 3	3 x Same as Study 1	Same as Study 1	<ul style="list-style-type: none"> • Cause (same as Study 1) • Praise/blame (revised Study 2) • Responsible (same as Study 2) • Punitiveness • Probability
Study 4	5 x Same as Study 1 Plus avalanche and flooding	2 Outcome (positive, negative) x 4 Distal event (positive action, negative action, abnormal physical, normal physical) x 2 Proximal event (abnormal physical, normal physical)	<ul style="list-style-type: none"> • Same as Study 3 • Abnormality
Study 5	7 x Fire, train, factory, avalanche, car, casino and soldier.	2 Outcome (positive, negative) x 3 Distal event (positive action, negative action, abnormal physical)	<ul style="list-style-type: none"> • Cause (same as Study 1) • Praise/blame (same as Study 3) • Probability (same as Study 3)
Study 6	7 x Same as Study 5	2 Outcome (positive, negative) x 3 Distal motive (positive action, negative action, unstated motive action)	<ul style="list-style-type: none"> • Same as Study 5 • Intent (2 items: belief & desire) • Intentionality • Foresight (2 items: foresight, predictability) • Unitisation of Chain (revised Study 2)

A second copy of Table 1.1 is available at the back of the Appendices for you to detach for your reference while reading this thesis.

Chapter Two: Study 1 - Human Agency, Motive and Outcome Information

Study 1 examined how attributions of cause and blame are influenced by the type of distal event (intentional action or physical event), the valence of the motive (positive or negative) behind the action, and the valence (positive or negative) of the outcome in opportunity chains (Hilton et al., 2005). Using scenarios with the same causal structure allowed me to first, test the conditions necessary to support the ‘special status’ that people ascribe to human agency over physical events as proposed by Hart and Honoré (1985).

The second goal was to test whether the causal relations between successive events are better accounted for by legal theories of causal reasoning (Hart & Honoré, 1985; Moore, 2009) or Heider’s (1958) conceptualisation of equifinal reasoning using the measure of unitisation of chain. This study also tests the predictions of both Tetlock et al.’s (2007) social functionalist approach and Alicke’s (2000) culpable control model. Both approaches predict main effects for outcome and type of distal event, where negatively valenced conditions will be rated higher than positively valenced conditions. The culpable control model, however, predicts that these main effects will be qualified by a three-way interaction between outcome, type of distal event, and temporal position, where people see negative actions, but not positive actions or physical events, as more causal when the outcome is negative rather than positive.

Method

Design and participants.

The study used a mixed design with three between-subject variables: distal event (a deliberate positively motivated action, a deliberate negatively motivated action⁷, or a physical event), outcome (positive or negative) and scenario (fire, train, or factory). A fourth within-subject variable was the temporal position of the causal event (distal or proximal). The proximal physical event was the same for all conditions in each scenario. The 3 (distal event) x 2 (outcome) design resulted in six versions of each scenario. Participants were voluntarily recruited on campus from Victoria University of Wellington ($N = 132$) and were randomly assigned to one of six between-subjects conditions. Participants were given a confectionary in appreciation of their participation. This procedure was replicated in Studies 2-6, so is not detailed again.

⁷ Deliberate positively motivated actions and deliberate negatively motivated actions are hereafter referred to as positive action or negative action respectively.

Scenario.

Each scenario depicted a sequence of two events leading to either a positive or negative outcome. See Appendix A, pp. 129-132. The first (distal) event had three conditions that were equivalent in terms of their effects on the outcome; a positively motivated action, a negatively motivated action, or a physical event. The voluntary action condition was described in two separate statements: an action statement and a motive statement. The action statement described the same action in both voluntary conditions that differed only in the agent's title, e.g., an arsonist, or a Conservation Department officer deliberately set fire to an area of forest. The motive statement described the agent's goal. The second (proximal) event was a physical event that was the same for all conditions in each scenario (e.g., strong wind). With the forest fire scenario, for example, the three distal event conditions leading to a positive outcome read:

Positive action: *A Conservation Department officer deliberately set fire to an area of forest. His goal was to cause a burn-off to prevent a larger fire from spreading. A strong wind developed which fanned the flames. These events prevented the larger fire from reaching the nearby town, saving dozens of homes.*

Negative action: *An arsonist deliberately set fire to an area of forest. His goal was to cause a larger fire to spread. A strong wind developed which fanned the flames. These events prevented the larger fire from reaching the nearby town, saving dozens of homes.*

Physical event: *A lightning strike set fire to an area of forest. A strong wind developed which fanned the flames. These events prevented the larger fire from reaching the nearby town, saving dozens of homes.*

For the three distal event conditions leading to a negative outcome, the scenarios were identical to those for positive outcomes, except that the last sentence read: *These events led to the larger fire reaching the nearby town, destroying dozens of homes.*

Measures and procedure.

Participants were given one of three scenarios (fire, train, factory) adapted from McClure et al. (2007). Each participant then completed a questionnaire that elicited judgments of the distal and proximal events' causality, blameworthiness, and unitisation of chain on a rating scale. Ratings of the voluntariness of each event were obtained as a manipulation check.

The measures of cause and blame replicated those used by Lagnado and Channon (2008). The measure of cause read: 'From the following statements, please rate the extent to which each event was the cause of the outcome in this scenario'. The stem was followed by the distal event e.g., The Conservation Department officer deliberately setting fire to the forest, and the second

physical event e.g., The strong winds. Ratings of each event were made on a scale, where 0 = Not at all the cause and 100 = Completely the cause. The measure of blame read: 'From the following statements, please rate the extent to which each event was to blame for the outcome in this scenario'. The stem was followed by rating scales for each event, as in the cause ratings. The anchors were: 0 = Not at to blame and 100 = Completely to blame.

The unitisation of chain measure replicated McClure et al.'s (2007) and Hilton and McClure (2013), and read: 'To what extent do you think the two events described above were two separate events or form part of a single chain of events? The stem was followed by a rating scale, where 0 = Two separate events, and 100 = A single chain of events. The measure of voluntariness replicated McClure et al. (2007) and read: 'From the following statements, please rate how voluntary you regard the action/event? The stem was followed by rating scales for each event, where 0 = not at all voluntary and 100 = completely voluntary.

Results

A mixed design ANOVA was conducted with three between-subjects variables: 3 (scenario: forest fire, train, factory) x 3 (distal event: positive action, negative action, physical event), x 2 (outcome: positive, negative) and two within-subjects variables: 2 (temporal position: distal, proximal) x 2 (judgment: cause, blame). Scenario did not interact with any other factors, so the analysis was repeated without this factor. See Figure 2.1 and Appendix B for descriptive statistics.

As a manipulation check to confirm that participants saw events designed to be intentional actions as more voluntary than events designed to be physical events, a 3 (distal event) x 2 (outcome) x 2 (temporal position) ANOVA was conducted for ratings of voluntariness (Appendix B). Intentional actions were seen as more voluntary than physical events as shown by a 2-way interaction between distal event and temporal position, $F(2, 108) = 23.06, p < .001, \eta^2 = .30$. Paired t -tests showed positive actions and negative actions were rated more voluntary than proximal physical events, $t(42) = 5.05, p < .001$, and $t(42) = 9.66, p < .001$ respectively. Distal physical events did not differ from proximal physical events. There were no interactions for outcome.

The 3 (distal event) x 2 (outcome) x 2 (temporal position) x 2 (judgment) mixed ANOVA revealed a main effect for judgments, Pillai's $V = 0.92, F(1, 126) = 12.50, p < .001, \eta^2 = .09$; ratings of cause ($M = 59.63, SE = 1.50$) were higher than ratings of blame ($M = 53.21, SE = 1.82$). There were no main effects for outcome, distal event, or temporal position, but importantly

there were two 3-way interactions⁸ The first interaction was between outcome, distal event and temporal position, Pillai's $V = 0.20$, $F(2, 126) = 12.88$, $p < .001$, $\eta^2 = .17$, and the second interaction was between distal event, temporal position and judgment, Pillai's $V = 0.07$, $F(2, 126) = 4.90$, $p < .01$, $\eta^2 = .07$.

Figure 2.1. Means of cause and blame for distal and proximal events by condition (Study 1).

Influence of distal event, temporal position, and outcome on attributions.

Several analyses clarified the first 3-way interaction between outcome, distal event and temporal position. To test predictions regarding the influence of the distal event on attributions of cause and blame, separate mixed design ANOVAs were conducted for chains with positive or negative outcomes. The ANOVAs were performed with between subjects variable: 3 (distal

⁸ When reading scenarios describing positive actions or positive outcomes, participants' responses might have been influenced by using a measure defined as blame rather than by the studies' main manipulations. This possibility is examined in Study 2.

event), and the within-subjects variables: 2 (temporal position) x 2 (judgments) for ratings of cause and blame.

With negative outcomes, there was a 2-way interaction between distal event and temporal position, $F(2, 65) = 17.77, p < .001, \eta^2 = .35$. Separate ANOVAs on distal and proximal events clarified this interaction. As predicted, the type of distal event influenced judgments of distal causes, $F(2, 65) = 11.55, p < .001, \eta^2 = .26$, but not exactly as predicted. Post hoc Bonferroni tests ($p < .05$) showed that negative actions were rated *higher* than distal physical events and positive actions, which did not differ. The type of distal event also influenced judgments of the proximal physical event, as shown by a main effect, $F(2, 65) = 11.55, p < .001, \eta^2 = .26$. Proximal physical events were rated lower when they followed negative actions than distal physical events, but they were rated higher when they followed positive actions than distal physical events. In contrast to negative outcomes, with positive outcomes there were no main effects for distal event, temporal position, or interactions. This indicates that cause and blame judgments did not differ for physical events or voluntary actions.

To test predictions regarding the influence of the valence of the outcome on attributions of cause and blame, separate mixed design ANOVAs were conducted comparing each type of distal event across positive and negative outcomes. The results show that negative outcomes amplify ratings for negative actions (cause, $F(1, 43) = 9.18, p < .001, \eta^2 = .17$; blame, $F(1, 43) = 5.56, p < .03, \eta^2 = .12$), but did not influence ratings for positive actions or distal physical events. In contrast when comparing the three proximal physical events across both outcomes, the results show that negative outcomes attenuate ratings for proximal physical events regardless of the type of preceding distal, (cause, $F(1, 43) = 11.59, p < .001, \eta^2 = .21$; blame, $F(1, 43) = 7.66, p < .01, \eta^2 = .15$).

Influence of distal event on attributions for consecutive events.

Correlational analyses clarified the 3-way interaction between the distal event, temporal position and judgment reported above. As predicted, the relations between distal and proximal events depended on the type of distal event, but the patterns differed for judgments of cause and blame ($p < .001$). For judgments of cause, there was a compensatory relationship between distal actions and proximal physical events: positive actions, $r = -.68$, and negative actions, $r = -.38$, were negatively related to proximal physical events, whereas in the baseline condition of two physical events, judgments for distal and proximal events were unrelated. For judgments of blame, a compensatory relationship was found between judgments for positive actions and

proximal physical events, $r = -.55$, but not between negative actions and proximal physical events or between distal and proximal physical events.

Influence of the unitary nature of the chain on attributions.

A univariate ANOVA was conducted with two between-subject variables: 3 (distal event) x 2 (outcome) for judgments of the unitisation of chain (see Table 2.1 for descriptive statistics). Counter to predictions, there were no main effects or interactions for outcome or type of distal event. Covariance analysis tested if the unitary nature of the chain influenced causal judgments using the difference score between the proximal and distal events. The analyses removed unitisation ratings from the interaction between outcome and distal event on causal ratings. The results showed that judgments of the unitisation of chain ($\beta = .21$) did influence causal judgments, $F(1, 122) = 7.47, p < .01$, partial $\eta^2 = .06$. However, after covarying out ratings of unitisation on judgments of cause, the interaction between outcome and distal event remained significant, $F(2, 122) = 12.28, p < .001$, partial $\eta^2 = .17$.

Table 2.1

Means, and SDs for Unitisation of the Chain Ratings by Condition (Study 1)

Variable	Positive Outcome	Negative Outcome
Positive Action	54.09 (38.50)	57.14 (36.21)
Negative Action	60.48 (38.53)	55.22 (39.30)
Physical Event	52.00 (39.93)	55.91 (40.43)

This suggests that how people see the relations between the events in the chain does influence their judgments of cause; people see integrated sequences of events as more causal than a series of separate coincidental events. However, the valence of the agent's motive and the outcome has the more influential effect when determining the cause of an outcome.

Discussion

This study shows that in these opportunity chains, the distal event and the outcome strongly influence judgments of cause and blame. The first key finding is that judgments about intentional actions and physical events are very different for positive and negative outcomes despite each event involving the same causal forces and having the same effect (e.g., a fire being lit). For chains with negative outcomes, the type of distal event has a major effect. Distal events are judged more causal and blameworthy when they are negative actions (e.g., an arsonist lighting a fire) than when they are physical events (e.g. lightning strike) or positive actions (e.g., a

Conservation officer lighting a burn-off). Moreover, positive actions are rated less causal, but no more blameworthy than physical events. For chains leading to positive outcomes however, the large differences between causes seen with negative outcomes disappear; the framing of the distal event makes almost no difference to judgments of cause or blame.

The finding that negative actions with negative outcomes receive the highest cause and blame ratings is consistent with earlier studies (Hilton et al., 2010; Lagnado & Channon, 2008; McClure et al., 2007). Furthermore, this finding supports predictions that people prefer voluntary actions to physical events as the causal explanation of the effect (Hart & Honore, 1985), and are consistent with Tetlock's (2002) and Alicke's (2000) models. The finding that judgments are amplified when negative actions lead to harmful rather than beneficial outcomes is also consistent with earlier research demonstrating outcome-valence asymmetries (Alicke, 1992; Bohnet, et al., 1988; Kahneman & Tversky, 1984). Furthermore, while both Tetlock's (2002) and Alicke's (2000) models predict that positive actions lead to lower attributions for negative outcomes, neither predicts that positive actions would be judged less causal than physical events. Indeed, the central tenet of both theories is that it is people's ability to control and manipulate the world around them that results in human actions being considered as more important causes than environmental factors.

When the chains included positive actions, or the intended outcome did not occur, a very different pattern of judgments emerged to that with negative actions and outcomes. Challenging the claim that people see intentional actions as more causal than physical events (Alicke, 2000; Hart & Honoré, 1985; Tetlock, 2002), when the outcome was positive, cause and blame judgments for intentional actions and distal physical events did not differ. Furthermore, when the action was positively motivated but harm eventuated, or the intended harm did not eventuate, the findings suggest that people discount voluntary actions in favour of physical events. For example, in the fire scenario, when the action of lighting a fire was motivated by a desire to save a town but instead dozens of homes were destroyed, or the action was motivated to cause harm but instead the town was saved, the action was judged less causal and blameworthy than the subsequent strong winds.

The finding that positive actions were judged no more causal with negative than positive outcomes is consistent with the Alicke's (2000) culpable control model but not Tetlock's (2002) social functionalist model. The social functionalist model predicts that cause ratings are amplified with negative and positive outcomes, as there is utility in identifying exemplary behaviour for future-oriented intervention. In contrast, the culpable control model predicts that when the agent

is judged to have good motives, negative evaluations formed in response to the harmful outcome ‘has nothing to work on’, so the agent is seen as no more causal when things go awry than when they go as intended.

These findings clarify the conditions necessary for an intentional action to be judged the primary cause in a sequence of events and answer several questions raised in Chapter 1. First, the results suggest that people’s good intentions do buffer them from blame when the outcome goes awry. Second, they suggest that people do recognise events as having the same causal effect despite differing mechanisms and goals. Third, they show that people judge an individual who intended to cause harm (e.g., the arsonist) less harshly when the anticipated harm didn’t occur.

A second novel finding is that causal ratings of distal intentional actions are complemented by the flip reverse ratings for proximal events, despite the fact that the proximal physical event is the same event across the various conditions. The results show a negative correlation for causal ratings between distal and proximal events when the chain consists of intentional actions and physical events, but not when the chain consists of two physical events. Thus, in chains with intentional actions and physical events, when the distal action is rated high, participants rate the proximal physical event lower, whereas when the distal action is rated low, participants rate the proximal physical event higher. The specific pattern of attributions across the distal and proximal events however, appears to be governed by the valence of the outcome and motive. It is as if something has to be held accountable for the outcome; so when the outcome is bad, negative actions are identified as the cause, but when the action is positively motivated, participants identify the proximal physical event as the cause. In contrast, the opposite pattern emerges when the outcome is positive; when the action is positively motivated, participants see that action as the cause, but when the action is negatively motivated, they identify the proximal physical event as the cause. This pattern was not replicated however for judgments of blame, where bad actions are blamed regardless, and bad outcomes amplify these ratings.

These shifts in judgments of the proximal events are not due to any difference in the event itself which is unchanged; so the shift is due to changes in the distal action. This pattern implies firstly, that in chains including an intentional action, when people see one event as the main determinant of an effect, they discount other events that are necessary for the effect (McClure, 1998) Second, the degree of congruence or ‘match’ between the intentional action’s motive and the outcome determines when people amplify or discount the agent’s causal contribution in favour of subsequent physical events. For example, when the intended outcome is realised, or ‘matched’ the agent’s motive (i.e., negative motive preceded a negative outcome; or positive

motive preceded a positive outcome), participants amplify the agent's causal contribution while discounting the proximal physical event. In contrast, when the outcome went awry or there was a 'mismatch' between motive and outcome (i.e., positive motive preceded a negative outcome; or negative motive preceded a positive outcome), participants discount the agent's causal contribution in favour of the proximal physical event. This compensatory relationship is described by Kelley's (1973) compensatory schema and Heider's (1958) hydraulic assumption. However, the novel feature here is that participants make compensatory judgments about two events in chains where both are necessary for the outcome (McClure & Hilton, 1998), whereas Kelley's model applied to two concurrent causes of an effect.

In contrast to chains comprising intentional actions and physical events, the findings for chains comprising two physical events provide an interesting comparison. First, the predicted influence of outcome valence was not found; a negative outcome did not amplify attributions and a positive outcome did not attenuate ratings of physical events. Second, consistent with McClure et al. (2007) and extending their research to opportunity chains with positive outcomes, the results show that people rate both events as equally causal and blameworthy. Finally, the findings challenge several theories regarding people's perceptions of the unitary nature of the causal chain. For example, the findings do not support Hart and Honore's (1985) hypothesis that when chains consist of two equivalent physical events, people judge the most recent event as the primary cause. Similarly, the findings are not consistent with Heider's (1958) claim that people view the sequence of events as part of a plan with intentional actions, but not with physical events. Instead, the results suggest that Heider was right to claim that people's perception of the integrated nature of the chain does influence attributions of cause, but wrong in the account of equifinal-multifinal as the reasoning processes. Instead, it appears that the agent's motive and the outcome play the critical role in determining judgments of cause.

In sum, this study replicates earlier research and supports several predictions when the sequence of events conforms to the legal paradigm of negative actions with negative outcomes (Alicke, 2000; Hart & Honore, 1985; Lagnado & Channon, 2008; McClure et al., 2007; Tetlock et al., 2007). However, extending the research on valence to multiple target events in sequences including positive actions and outcomes goes some way to clarifying the anomalies reported in earlier research (Hilton & McClure, 2010; Hitchcock & Knobe, 2009; Lagnado & Channon, 2008; McClure et al., 2007). These findings show that ratings of positive actions and proximal events for negative outcomes are the flip reverse of those for negative actions and proximal

causes, despite the fact that the proximal event is always a physical cause that is constant across the various conditions. Yet with positive outcomes all causes are equal, or nearly so.

These findings suggest that regardless of outcome, lay reasoning starts from the premise of equality between two similar types of events, as shown by the equivalent ratings for distal and proximal physical events. However, when one event in the chain is a voluntary action, that action appears to dominate people's reasoning, or as Heider (1958) suggested, 'it seems that action has such salient properties it tends to engulf the total field rather than be confined to its proper position' (p. 54). Specifically, as people assess the core features of intentionality, motive, proximity, and outcome information, their judgments of cause and blame are moderated as each event's relative contribution to the outcome is apportioned against alternative events. This suggests that both legal and psychological theorists are right to emphasize the 'special status' attributed to voluntary actions but wrong in their understanding of causal relations and the influence of valence on lay reasoning.

Finally, it appears that the interaction between outcome valence, type of distal event and temporal position for judgments of cause and blame found in this study is best explained by Alicke's (2000) culpable control model. Specifically, the results obtained here challenge the premise that social controllability best explains judgments of cause (Tetlock, 2002). Instead, they support the claim that people's evaluative reaction to the agent, the events, and the outcome encourages them to interpret the evidence in a manner that supports their 'intuitive blame hypothesis' (Alicke, 2000). It is as if people are intuitively balancing the scales of justice.

Chapter Three: Explaining the Effects: Choice of Measures or Outcome Valence?

Study 1 shows that people's attributions are influenced by the valence of the agent's motive and the outcome. In chains leading to negative outcomes, the type of distal event affects judgments about both causes in the chain, in that positive and negative actions in the chain are judged very differently to each other and to physical causes. In contrast, when chains lead to positive outcomes, the type of distal event makes almost no difference to judgments of either the distal or proximal event. Despite these findings, some interesting questions arise. Are these effects an artefact of the measures used? For example, do the effects reflect linguistic rather than theoretical distinctions between the concepts of cause and blame? Alternatively, are these findings an artefact of the scenarios chosen? Specifically, do the effects reflect differences in the event's efficacy to bring about the intended outcome (Lombrozo, 2010)? For example, could the effects be explained by people seeing arsonists as more competent than Conservation officers at starting fires? Or could the outcome valence effects have been exaggerated by differences in the salience or severity of the scenario's outcome (McClure et al., 1991)? Could this property account for the effects on judgments of cause? Or are these findings a fluke? My interpretation of the results for Study 1 would be strengthened if those findings are replicated using more robust measures.

Study 2: Extending the Measures of Praise/Blame and Responsibility

Study 2 was designed as a stronger test of the predictions discussed in Chapter 2 while addressing several questions. The first question regards the measure of blame used in Study 1. The ascription of blame presupposes a harmful outcome, carries negative emotional connotations and implies that the blameworthy person deserves some form of punishment (Tetlock et al., 2007; Weiner, 1995). The word 'blame' is not usually used for positive actions. Thus, when reading scenarios describing positive actions or positive outcomes, participants' responses might have been influenced by using a measure defined as blame rather than by the studies' main manipulations. To enable judgments to better reflect the positive and negative valence of the actions in addition to physical events, the blame measure used in Study 1 was replaced with Alicke et al.'s (2011) bipolar measure of praise/blame in Study 2. Furthermore, using a praise/blame scale helps ensure participants are not misled as a result of thinking that there is some particular point in the provision of a blame question that would not normally be applied to physical events or positive conditions (Grice, 1975; Hilton, 1990, 1995).

The second question also concerns the measure of blame. Judgments of cause and blame are different constructs theoretically, yet lay meanings for these constructs are similar and often transposed. For example, people often ‘blame’ the weather for the cancellation of a cricket match rather than citing the weather as the cause of the cancellation. Therefore, it is possible that the effects reported in Study 1 may reflect linguistic rather than psychological differences (Alicke, 2008; McClure & Hilton, 1997). To address this issue, I added a measure of responsibility for two reasons. First, in contrast to the negative connotations associated with judgments of blame, responsibility is a more neutral assessment that can reflect people’s view of accountability for either positive or negative outcomes. Research has shown that people assign more blame for bad outcomes and actions than good ones (Alicke et al., 2008; Alicke et al., 1996), whereas their responsibility judgments distinguish less between good and bad outcomes (Enzle & Hawkins, 1992; Malle, 2006b). Thus, the valence of the outcome should influence blame judgments more than people’s responsibility judgments.

Second, including a responsibility measure enabled me to compare the distinct roles that intent, voluntariness, and outcome information play in determining judgments of cause, blame and responsibility. Research suggests that people distinguish between judgments of cause, blame and responsibility by applying different weightings to the central features (e.g., intentionality, control, foresight, knowledge, etc.) of the causal chain. Research has shown that intent, foreseeability and epistemic state affected judgments of blame more than cause, although the distinctions between responsibility and blame were less clear (Fincham & Jaspars, 1979; Hilton & McClure, 2013; Lagnado & Channon, 2008). For example, Lagnado and Channon (2008) reported an interaction between intention and foresight, where objective foresight had a stronger effect on judgments of cause, while subjective foresight had a stronger effect on judgments of blame⁹. Similarly, Cushman (2008) found that the agent’s intent had the stronger effect on judgments of permissibility or wrongness, whereas judgments of blame and punishment reflected both intent and outcome information.

Based on valence research (e.g., Alicke, 1992; Bohnet et al., 1988), I predicted that the valence of the outcome would influence ratings depending on the type of event (voluntary actions or physical events) and the type of judgment. Specifically, I predicted that ratings for voluntary actions would be higher for negative outcomes than positive outcomes, but the size of this effect

⁹ Objective foresight refers to how likely the outcome was from a factual point of view, regardless of what the agent believed. In contrast, subjective foresight refers to how likely the outcome was from the agent’s point of view.

would differ depending on the type of judgment; the effect would be greater for cause and blame ratings than responsibility ratings. In contrast, because the physical events in these scenarios are uncontrollable occurrences, people's responses are more likely to be influenced by the outcome of the sequence than the physical event itself. Therefore, I also predicted that physical events would be rated higher for negative outcomes than positive outcomes for judgments of cause but not for judgments of responsibility or blame.

To differentiate between the constructs of cause, blame and responsibility¹⁰ and test the claim that the construct of blame carries emotional connotations as well as assessments of liability (e.g., Alicke, 2000; Hart & Honoré, 1985; Weiner, 1995), I predicted that regardless of the outcome, ratings of blame would be higher for negative than positive actions, but that ratings of blame would be higher than responsibility ratings for negative actions and lower than responsibility ratings for positive actions. Thus, in statistical terms I predicted a 3-way interaction between the valence of the outcome, the type of distal event, and the type of judgment (cause, praise/blame and responsibility) for ratings of the distal event.

Method

Design and scenarios.

The design replicates Study 1. Three changes were made to the scenarios in Study 1 to reduce ambiguities. To emphasize the temporal position of the events in the scenarios, the phrase '*that night*' or '*at the same time*' was incorporated into the description of the proximal event. Second, in the fire scenario, to reduce possible ambiguity, the word '*wildfire*' was used to differentiate between the deliberately lit fire and the naturally occurring '*larger fire*'. For example, the distal positive action leading to a positive outcome read: *A Conservation Department Officer deliberately set fire to an area of forest. His goal was to cause a burn-off to prevent a wildfire from spreading. That night a strong wind developed which fanned the flames. These events prevented the wildfire reaching the nearby town, saving dozens of homes.*

Finally, in the factory scenario the phrase 'a disgruntled employee' was replaced with 'a vindictive employee' to reduce any imbalance in the positive and negative phrasing 'a diligent employee' and 'a disgruntled employee'.

Measures and procedure.

Participants ($N = 262$) read a single scenario then gave judgments of the distal and proximal events' causality, responsibility, praiseworthiness or blameworthiness, and the

¹⁰ See Chapter 1, p. 4 for a discussion (Hart & Honoré, 1985).

unitisation of chain on a rating scale. The measures of cause and unitisation were the same as Study 1. The manipulation check measure for voluntariness of each event was not repeated in Study 2. The measure of blame used in Study 1 was replaced with Alicke et al.'s (2011) evaluative measure, which read 'How praiseworthy or blameworthy do you think the following events were for the outcome in this scenario?' The stem was followed by a bipolar scale for each event, where 0 = Extremely praiseworthy, and 10 = Extremely blameworthy. The measure of responsibility was based on that used by Alicke (1992) and Alicke et al. (2008) and read: 'From the following statements, please rate the events in terms of their responsibility for the outcome in this scenario'. The stem was followed by rating scales for each event, where 0 = Not at all responsible and 10 = Completely responsible.

Dependent variables were converted using the linear transformation POMP (percent of maximum possible) method of scoring (Cohen, Cohen, Aiken, & West, 1999). This conversion of the dependent variables was replicated in Studies 3-6, so is not detailed again.

Results

To test whether the effects in Study 1 were replicated in Study 2, a mixed design ANOVA was conducted with three between-subjects variables: 3 (scenario: forest fire, train, factory) x 3 (distal event: positive action, negative action, physical event), x 2 (outcome: positive, negative) and two within-subjects variables: 2 (temporal position: distal, proximal) x 3 (judgment: cause, praise/blame, responsibility). Scenario did not interact with any other factors, so the analysis was repeated without this factor. See Figure 3.1 and Appendix C1 for descriptive statistics.

The analyses revealed main effects for outcome, $F(1, 253) = 14.23, p < .001, \eta^2 = .06$, and distal event, $F(2, 253) = 15.55, p < .001, \eta^2 = .11$, but not for temporal position, or judgment. The main effects were qualified by three 3-way interactions. Importantly, the first interaction was between outcome, distal event, and temporal position, $F(2, 253) = 3.42, p < .05, \eta^2 = .04$. The second interaction was between outcome, temporal position, and judgment, $F(2, 252) = 8.00, p < .001, \eta^2 = .06$, and the third interaction was between distal event, temporal position and judgment, $F(4, 506) = 10.99, p < .001, \eta^2 = .08$.

Figure 3.1. Means of cause, responsibility, and praise/blame for distal and proximal events by condition (Study 2).

Influence of distal event, temporal position, and outcome on attributions.

To clarify the interaction between outcomes, distal event, and temporal position, and to test predictions regarding the influence of the type of distal event on judgments of cause, praise/blame, and responsibility, separate mixed design ANOVAs were conducted for chains with positive or negative outcomes, with one between-subjects variable: 3 (distal event), and two within-subjects variables: 2 (temporal position) x 3 (judgment). See Appendix C2.

With negative outcomes, there was a 2-way interaction between the distal event and temporal position, $F(2, 125) = 17.44, p < .001, \eta^2 = .21$. Replicating Study 1 and supporting predictions, follow-up ANOVAs showed that the type of distal event influenced judgments of distal events, $F(2, 129) = 36.53, p < .001, \eta^2 = .36$. Negative actions were rated *higher* than distal physical events, which in turn were rated *higher* than positive actions. Unlike Study 1, the type of distal event did not influence judgments of proximal events.

With positive outcomes, unlike Study 1, there was a 2-way interaction between distal event and temporal position, $F(2, 125) = 3.56, p < .04, \eta^2 = .05$. Follow-up ANOVAs showed that the type of distal event influenced judgments of distal events, $F(2, 125) = 5.33, p < .01, \eta^2 = .08$, but not proximal events. Positive actions were rated *lower* than negative actions and distal physical events, which did not differ.

To test predictions regarding the influence of the valence of the outcome on judgments, separate mixed design ANOVAs were conducted comparing each type of distal event across positive and negative outcomes. Negative outcomes amplified ratings for negative actions, $F(1, 85) = 6.58, p < .001, \eta^2 = .08$, attenuated ratings for positive actions, $F(1, 85) = 7.87, p < .01, \eta^2 = .09$, but had no effect on ratings for distal physical events. In contrast, negative outcomes amplified ratings for proximal physical events regardless of the type of preceding distal event (positive actions, $F(1, 85) = 9.13, p < .01, \eta^2 = .10$; negative actions, $F(1, 85) = 6.16, p < .01, \eta^2 = .07$; physical events, $F(1, 84) = 7.47, p < .01, \eta^2 = .08$).

Influence of distal event on attributions for consecutive events.

Correlational analyses clarified the relations between distal and proximal events for judgments of cause, praise/blame and responsibility. Replicating Study 1, these relations reflected the type of distal event. For chains comprising voluntary actions and physical events, there was a compensatory relationship between judgments of distal actions and proximal physical events (positive actions: cause, $r = -.57, p < .001$; praise/blame, $r = -.37, p < .001$; responsibility, $r = -.57, p < .001$; and negative actions: cause, $r = -.38, p < .001$; praise/blame, $r = -.25, p < .02$; responsibility, $r = -.38, p < .001$). In contrast, in the baseline condition of chains with two

physical events, the relations between events differed depending on the type of judgment. Distal and proximal events were negatively related for ratings of cause, $r = -.36, p < .001$, and praise/blame, $r = -.25, p < .02$, but not related for responsibility.

The relations between judgments of cause, responsibility and praise/blame.

Correlational analyses examined the relations between these three types of judgments using the difference score between distal and proximal events (Appendix C3). As predicted, in chains with negative outcomes there were positive relations between all judgments (cause and responsibility, $r = .67, p < .001$; cause and praise/blame, $r = .61, p < .001$; praise/blame and responsibility, $r = .67, p < .001$). With positive outcomes, a different pattern emerged; judgments of cause and responsibility were positively related ($r = .53, p < .001$), but not cause and praise/blame, or praise/blame and responsibility.

In sum, the two 3-way interactions with judgment reported above show that judgments of cause, praise/blame, and responsibility differ depending on voluntariness (shown by the temporal order effects), the type of distal event, and the valence of the outcome, but not exactly as predicted. I had predicted that people distinguish between judgments of cause, blame, and responsibility by applying different weightings to voluntariness, motive and outcome valence. Notwithstanding these distinctions, I proposed that judgments of responsibility would show a similar pattern to judgments of cause, both of which would be very different to judgments of blame. Instead, follow-up analyses showed that judgments of responsibility sometimes function like judgments of cause, and sometimes like judgments of blame (see Appendix D for analyses).

Influence of the unitary nature of the chain on attributions.

To test whether the valence of the agent's motive or outcome influenced judgments of unitisation, an ANOVA was conducted with two between-subject variables: 3 (distal event) x 2 (outcome) for judgments of unitisation. See Table 3.1 for descriptive statistics.

Table 3.1

Mean and SDs for Unitisation of Chain by Condition (Study 2)

<i>Distal Event-Type</i>	Positive Outcome	Negative Outcome
Positive Action	43.10 (34.20)	39.30 (33.85)
Negative Action	50.90 (37.91)	58.90 (30.29)
Physical Event	45.40 (33.49)	60.70 (35.57)

There was a main effect for distal event, $F(2, 244) = 4.34, p < .02$, where chains with positive actions were seen as two separate events more than chains with negative actions or two physical events, which they saw more as a single causal chain. There was no main effect for outcome or interaction between outcome and distal event. As in Study 1, the failure to find these effects for differing types of distal event eliminates unitisation as a plausible explanation of causal attributions in these chains.

Discussion

The results from Study 2 replicate those reported in Study 1 for voluntary actions with negative outcomes. As predicted, in chains with negative outcomes, negative actions (e.g., an arsonist lighting a fire) are seen as more causal, blameworthy and responsible for harmful consequences than positive actions (e.g., a Conservation officer lighting a fire), even when these events have an equivalent causal effect and involve the same core features of intent, control, and proximity. In chains with positive outcomes, the results from Study 2 replicates Study 1 for judgments of cause, but not blame: the valence of the action has little effect on ratings of cause, but a significant effect on ratings of blame. In these scenarios, negative actions are blamed, but positive actions do not appear to sway the moral pendulum.

As in Study 1, there was mixed support for predictions comparing judgments for voluntary actions and physical events. For chains with negative outcomes, the valence of the agent's motive has a major effect on judgments: distal physical events (e.g., a lightning strike) with the same effect on the outcome as the intentional actions are seen as less causal, responsible or blameworthy than negative actions, but more causal, responsible and blameworthy than positive actions. For chains with positive outcomes, the valence of the motive had little effect on judgments of cause or responsibility but did influence judgments of blame.

Judgments for proximal physical events in Study 2 differed from those in Study 1. In Study 2, the valence of the outcome rather than the type of distal event or the motive for preceding actions influenced judgments for physical events. For example, when the outcome was negative, these results replicate McClure et al.'s (2007) findings that people see consecutive physical events as being equally causal. But contrary to predictions, with positive outcomes, participants discounted the causal contribution of proximal physical events in favour of distal physical events. Indeed, it appears that by emphasising the temporal order of the events by adding the phrase for example, *'that night'* or *'at the same time'*, in Study 2 reduced ambiguity that may have influenced results in Study 1.

The patterns in Study 2 replicate the effects of the valence of the agent's motive on ratings of blame in Study 1 even when using a bipolar praise/blame rating scale. Notably, physical events were blamed and rated higher than positive actions, which in turn were judged, albeit modestly, praiseworthy. This suggests that the effects reported for physical events, positive actions or positive outcomes in Study 1 were not confounded by violating the communicative principles of relevance (Grice, 1975).

Study 2 also shows that people see judgments of cause, blame and responsibility as closely related, yet distinct constructs. When people assess positive or negative actions, they distinguish between the constructs of praise/blame and responsibility, but not between cause and responsibility. As predicted, the direction of differentiation between praise/blame and causal responsibility also diverged, depending on the valence of the action. In contrast to the distinctions between constructs made for voluntary actions, with physical events, ratings for responsibility differed to those for cause and blame.

These results suggest that people's judgments of praise or blame reflect the morality of the agent's action, whereas judgments of cause and responsibility reflect the type of event (voluntary or physical). The finding that judgments of cause and responsibility were distinguished between in the baseline condition of chains comprising two physical events but not in chains comprising voluntary actions is interesting for several reasons. First, it suggests that people do recognise the semantic differences between the constructs but they do not apply those distinctions as proposed by legal theorists (Hart & Honoré, 1985). Secondly, they suggest that responsibility is seen as a relatively objective term that is largely unaffected by voluntariness, intent, or morality. The finding that physical events are blamed challenges the proposal that blame, being a moral judgment, can only be ascribed to voluntary actions (Hart & Honoré, 1985; Knobe, 2010). However, the finding that participants did not distinguish between cause and praise/blame when the chains only comprised physical events, suggests that people do use the term 'blame' to describe a causal relationship in the absence of voluntary actions.

Overall, these findings add to earlier research on the effects of valence on different types of judgments (Fincham & Jaspars, 1979; Hilton & McClure, 2013; Lagnado & Channon, 2008), and suggest that negativity amplifies judgments of cause and blame but not responsibility. This supports the claim that blame carries emotional connotations as well as assessments of liability, whereas judgments of responsibility are seen as more neutral assessments of accountability (e.g., Weiner, 1995).

Study 3: Adding Measures of Punitiveness and Probability

Study 3 was designed as a stronger test of the valence effects reported in Study 2 while addressing three questions. The first question concerns the measures of cause and responsibility. In Study 2, the predicted distinctions between these two constructs did not emerge. Instead, the effects for judgments of cause and responsibility were almost identical when the chain included voluntary actions but not with physical events. It is possible that the lack of differentiation between the constructs of cause and responsibility reflects participants' interpretation of the questions being asked and the inferred explanation being sought rather than theoretical definitions (Hart & Honoré, 1985; Hilton, 1995).

The second question examines whether the effects of voluntariness, motives and outcomes on people's judgments of cause, blame and responsibility, differ for judgments of reward or retribution. Legal theory suggests that judgments of punitiveness should resemble judgments of blame (Hart and Honoré, 1985; Moore, 2009). Yet Tetlock et al. (2007) suggested that the differing effects of voluntariness and morality on judgments of cause and responsibility reflect people's perception of the social utility of identifying a target event for future intervention. Indeed, using scenarios where the defendant causes minor or severe harm, Tetlock et al. found that measures of punitiveness were more strongly correlated with judgments of responsibility than judgments of cause or blame. I added a bipolar measure of punitiveness to address these two questions.

The findings in Study 2 for negative outcomes support Tetlock et al.'s (2007) claim when we compare the results for chains with voluntary actions against physical events in the baseline condition. Participants' attributions in the baseline condition mirrored those for chains with negative actions, but their judgments of responsibility were the same as chains with positive actions. Thus, it is possible that when events lead to harm, people judge negative actions and distal physical events as the primary cause because they are more appropriate targets for intervention or social control. However, as physical events in themselves are uncontrollable occurrences, these events cannot be held liable for the outcome, thus mitigating judgments of responsibility. Adding a measure of punitiveness allowed me to firstly, differentiate between the constructs of cause and responsibility, and secondly, to evaluate whether the patterns of judgments are better accounted for by the culpable control model (Alicke, 2000), social functionalist theory (Tetlock, 2002), or legal theory (Moore, 2009).

Based on Tetlock et al.'s (2007) research, I extended the predictions in Study 2 to include judgments of punitiveness. I predicted that the positive correlation between responsibility and punitiveness would generalise to chains with positive actions and outcomes. Specifically, I predicted that in chains with negative outcomes, negative actions would be rated more causal, blameworthy, responsible and deserving of punishment than positive actions. Similarly, in chains with positive outcomes, negative actions would also be rated more blameworthy and deserving of punishment than positive actions, but this effect would not be seen with judgments of cause or responsibility. Second, based on legal claims that inchoate crimes are half as blameworthy as realised crimes (Moore, 2009), I predicted that when agents fail in their attempt to cause harm (i.e., when negative actions precede positive outcomes), their actions would be rated less causal, responsible, blameworthy and deserving of punishment than agents who succeeded in causing the harmful outcome. Third, I predicted that positive actions would be rated more responsible and deserving of punishment with negative than positive outcomes, but not with judgments of cause or praise/blame. Thus, in statistical terms, I predicted a 3-way interaction between the valence of the outcome, the type of distal event, and judgments. Furthermore, I predicted that judgments of causality would be best explained by an index of social control over voluntary and physical events (measured by judgments of punitiveness).

The third question I addressed in Study 3 concerns the scenarios. First, I was interested in whether the differences in the perceived efficacy of an event to cause the intended outcome rather than voluntariness *per se* could explain the effects on judgments of causality (Lombrozo, 2010). For example, people may perceive a 'man lighting a fire' to be a more reliable method of starting a forest fire than a 'lightning strike'. Second, I was thought it was possible that the valence effects may have been exaggerated by differences in the severity of the outcome (McClure et al., 1991; Tetlock et al., 2010). For example, people might not see 'houses being saved' as the opposite of 'houses being destroyed'. Indeed, participants may have seen houses being saved as the normal condition, or the absence of an effect, but houses being destroyed may have been seen as a vivid, abnormal occurrence. Research has shown that people overweight improbable events (Kahneman & Tversky, 1982), engage in more causal reasoning when the event is unexpected, abnormal, or inconsistent with available schemata (Bohner et al., 1988; Kahneman, Slovic, & Tversky, 1982). Alternatively, the evaluative response elicited by negative actions or outcomes may prompt participants to reason counterfactually about what might have been. The extent the events then conformed to social or functional norms may have mediated attributions (Epstude & Roese, 2008; Roese, 1997).

To examine these issues, I added Mandel's (2003) measure of outcome probability. I predicted that when the type of event was controlled and the positive and negative versions of the outcome described in the scenario are diametrically opposed, the valence of the outcome should not have an effect on judgments of the likelihood the outcome would occur. With the fire scenario, for example, when the sequence of two physical events results in a forest fire, the consequences (e.g., the houses being saved or destroyed) should make no difference to the probabilistic contribution of each event to the outcome.

In addition, the measure of outcome probability allowed me to test whether causal attributions might be explained by probability (e.g., Spellman, 1997). Specifically, as noted in Chapter 1, McClure et al. (2007) found that participants' judgments of each event's probabilistic contribution to negative outcomes did not predict judgments of cause. However, it is possible that causal attributions reflect a degree of foreseeability of the outcome rather than probabilistic increase *per se*. Thus, people might infer that in acting intentionally, the agent possesses some further information that might facilitate the desired outcome. For example, participants might assume that the Conservation officer would have taken all reasonable precautions and checked weather conditions before lighting a burn-off, thus mitigating his causality when the outcome goes wrong. In contrast, participants might assume that an arsonist would either ignore or contrive to take advantage of the weather conditions, thus amplifying attributions of causality. Spellman's (1997) updating probability model accounts for moral valence effects based on the proposal that attributions of cause are conditional on what has gone before and the observer's causal inferences and world knowledge. This theory predicts that people attribute causation to the event that most increases the probability of the outcome and that indices of probabilistic contribution (the probability of the outcome given the presence or absence of the preceding events) mediate the effects of the type of event (voluntary or physical) on judgments of cause. Based on McClure et al.'s findings, I expected less support for this model than functionalist approaches.

To summarise, I added the measures of punitiveness and probability to the measures used in Study 2 for two reasons: first, to examine the ability of the social functionalist and probabilistic approaches to predict judgments of cause, and second, to clarify the relations between the judgments of cause and responsibility. The measure of unitisation of chain was eliminated because it was unable to explain the effects in Study 1 or 2.

Method

Design and measures.

The design and scenarios replicate Study 2. Participants ($N = 215$) read a single scenario then gave judgments of the distal and proximal events' causality, responsibility, praiseworthiness or blameworthiness, and punitiveness on a rating scale. The measures of cause and responsibility were the same as in Study 2. The anchors on the praise/blame measure were revised to accentuate the presence an effect, and the neutral midpoint was explicitly named to better accommodate physical events. Thus, the praise/blame stem was followed by a bipolar rating scale for each event, where -5 = Extremely blameworthy, 5 = Extremely praiseworthy, and 0 = Neither.

The measure of probability replicated McClure et al.'s (2007, Expt. 4) adaptation of Mandel's (2003, Expt. 2) probability task. The instruction read: 'From the following statements, please rate your estimate of the **probability** of the outcome in this scenario (*e.g. the houses being saved*) for the three following scenarios: e.g., (1) Given that the Conservation Department Officer **did not** set fire to an area of forest and the strong winds **did not** fan the flames; (2) Given that the Conservation Department Officer **did** set fire to an area of forest but the strong winds **did not** fan the flames; (3) Given that the Conservation Department Officer **did** set fire to an area of forest and the strong winds **did** fan the flames.' Judgments were made on a scale from 0% to 100%, where 0% = No chance and 100% = Absolutely certain.

The measure of punitiveness was based on Tetlock et al.'s (2007) punitiveness-toward-individual scale which assessed the degree to which the agent should be subject to the maximum disciplinary penalties. To accommodate positive actions and outcomes, the measure was adapted to include a bipolar reward/punishment rating scale. The measure read: 'How much do you think society should give **rewards or punishments** in response to events like these?' The stem was followed by a bipolar scale for each event, where -5 = Maximum punishment allowed, 5 = Maximum reward available, and 0 = Neither, represented the neutral midpoint position.

The measures of reward/punishment and praise/blame were reverse coded to match the valence of the anchors in Study 2 and 3, where 0% = Extremely praiseworthy or Maximum reward allowed, and 100% = Extremely blameworthy or Maximum punishment allowed. The measures of cause, responsibility, praise/blame and punitiveness were counterbalanced in a Latin square design, and the order of probability attributions was reversed for half of the conditions. A MANOVA revealed no effects involving the order of dependent variables, so subsequent analyses were collapsed for order and are not discussed further.

Results and Discussion

An omnibus mixed ANOVA was conducted with three between-subjects variables: 3 (scenario: forest fire, train, factory) x 3 (distal event: positive action, negative action, physical event), x 2 (outcome: positive, negative) and two within-subjects variables: 2 (temporal position: distal, proximal) x 4 (judgment: cause, responsibility, praise/blame, punitiveness). Scenario did not interact with any other factors, so subsequent analyses were repeated without this factor. Descriptive statistics are shown in Figure 3.2 and Appendix E1.

As in Study 2, there was a main effect for outcome, $F(1, 209) = 21.10, p < .001, \eta^2 = .09$, and type of distal event, $F(2, 209) = 13.94, p < .001, \eta^2 = .12$. There were no main effects for temporal position or judgments. Importantly, there was a 4-way interaction between outcome, distal event, temporal position, and judgment, $F(6, 416) = 6.66, p < .001, \eta^2 = .09$.

The relations between cause, responsibility, praise/blame and punitiveness.

Correlational analyses examined the relations between the four judgments using the difference score between distal and proximal events for positive and negative outcomes (see Table 3.2 and Appendix E2).

Table 3.2

Correlations Between Judgments Using the Difference Score Between Ratings for Distal and Proximal Events (Study 3)

	<i>Cause</i>		<i>Responsibility</i>		<i>Praise/blame</i>	
	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome
<i>Cause</i>	-	-				
<i>Responsibility</i>	.51***	.72***	-	-		
<i>Praise/Blame</i>	-.07	.65***	-.09	.59***	-	-
<i>Punitiveness</i>	-.22*	.43***	-.06	.47***	.76***	.61***

* $p < .05$; ** $p < .01$; *** $p < .001$.

Replicating Study 2, with negative outcomes, the analyses showed positive relations between all judgments. With positive outcomes, as in Study 2, the analyses also revealed different relations between judgments. Judgments of cause were positively related to judgments of responsibility and punitiveness, but not praise/blame. Judgments of praise/blame however, were positively related to punitiveness. However, the prediction that the positive relations between judgments of responsibility and punitiveness would generalise to positive outcomes was

Figure 3.2. Mean ratings of cause, responsibility, praise/blame and punitiveness for distal and proximal events by conditions (Study 3).

not supported. These findings show regardless of outcome, people see judgments of cause and responsibility, and judgments of praise/blame and punitiveness as two distinct groups of constructs.

The finding that judgments of cause related to judgments of punitiveness for both positive and negative outcomes (the diverging direction reflects the bipolar punishment/reward scale) supports Tetlock et al.'s (2007) claim that people recognise events as causes of good or bad outcomes for the purpose of retribution or future intervention. However, these findings do not support Tetlock et al.'s claim that judgments of responsibility and punitiveness are 'almost identical constructs when harm eventuates' (p. 199).

Influence of distal event, temporal position, and outcome on attributions.

Targeted analyses for distal events examined firstly, whether the distinct effects of voluntariness, motive and outcome information on people's judgments of cause, blame and responsibility were replicated in Study 3, and secondly, whether these effects differ for judgments of punitiveness, as indicated by the 4-way interaction reported above. Separate MANOVAs were performed with 3 (distal events) for chains with positive or negative outcomes for ratings of the distal event for the four judgments. There was a multivariate main effect for distal event for chains with positive outcomes, $F(8, 206) = 8.51, p < .001, \eta^2 = .25$, and negative outcomes, $F(8, 204) = 8.01, p < .001, \eta^2 = .24$. Follow-up univariate ANOVAs showed judgments differed depending on the type of distal event (see Table 3.3 for inferential statistics).

Supporting predictions, in chains with negative outcomes, negative actions were rated more causal, responsible, blameworthy, and deserving of punishment than positive actions. In contrast and contrary to predictions, in chains with positive outcomes, negative actions were rated more blameworthy and deserving of punishment but no more causal or responsible than positive actions. For distal physical events, judgments of cause and responsibility differed from voluntary actions with negative outcomes but not positive outcomes. With negative outcomes, distal physical events were rated as causal as negative actions, but as responsible, blameworthy, and deserving of punishment as positive actions.

To test predictions regarding the influence of the valence of the outcome on attributions, separate MANOVAs were conducted comparing each type of distal event across positive and negative outcomes. The results show that the valence of the outcome influenced ratings of distal events, as shown by multivariate main effects for chains comprising positive actions, $F(8, 63) = 7.27, p < .001, \eta^2 = .48$, negative actions, $F(8, 62) = 4.23, p < .001, \eta^2 = .35$, and physical events, $F(8, 63) = 5.03, p < .001, \eta^2 = .39$. Follow-up univariate analyses showed negative outcomes

Table 3.3
Summary of Univariate ANOVAs for Distal Judgments of Cause, Praise/Blame, Responsibility and Punitiveness (Study 3)

Variable	Type of Distal Event				Valence of Outcome				
	<i>df</i>	<i>F</i>	<i>p</i>	η^2		<i>df</i>	<i>F</i>	<i>p</i>	η^2
Cause									
Positive Outcome	(2,104)	0.90	ns		Positive Action	(1, 70)	9.00	.01	.11
Negative Outcome	(2,105)	12.95	.001	.20	Negative Action	(1, 69)	5.61	.02	.08
					Physical event	(1, 70)	0.15	ns	
Responsibility									
Positive Outcome	(2,104)	1.41	ns		Positive Action	(1, 70)	13.80	.001	.17
Negative Outcome	(2,105)	11.26	.001	.18	Negative Action	(1, 69)	3.09	ns	
					Physical event	(1, 70)	0.22	ns	
Praise/Blame									
Positive Outcome	(2,104)	29.57	.001	.36	Positive Action	(1, 70)	8.66	.01	.11
Negative Outcome	(2,105)	24.42	.001	.32	Negative Action	(1, 69)	1.02	ns	
					Physical event	(1, 70)	4.33	.05	.06
Punitiveness									
Positive Outcome	(2,104)	36.64	.001	.41	Positive Action	(1, 70)	12.83	.001	.16
Negative Outcome	(2,105)	21.26	.001	.29	Negative Action	(1, 69)	0.13	ns	
					Physical event	(1, 70)	0.61	ns	

amplified judgments of cause, but not for blame, punitiveness, or responsibility for negative actions. In contrast, negative outcomes attenuated judgments of praise, reward, cause and responsibility. For ratings for distal physical events, the valence of the outcome did not influence judgments of responsibility or punitiveness, but did influence judgments of blame: negative outcomes amplified ratings of blame. Contrary to predictions, there were no outcome effects for ratings of cause.

In sum, as shown in Figure 3.2 and the analyses reported in Table 3.3, judgments of cause, praise/blame and responsibility replicate those found for distal events in Study 2. Judgments of cause and responsibility are influenced by the agent's motive and the consequences, whereas judgments of punitiveness and praise/blame were influenced by the valence of the agent's motive.

Overall, these findings show when things end well, what, why or when an event occurred makes little difference to people's attributions of cause and responsibility; all things are equal. But when there are harmful consequences, the morality of the action and the valence of the outcome determine judgments of cause and responsibility. In contrast, regardless of the outcome, the morality of the agent's action has a strong effect on judgments of praise/blame and punitiveness. Counter to the legal claim that a guilty mind is not necessary for the ascription of responsibility, these findings suggest that responsibility requires a causal connection to the outcome, but a guilty mind is sufficient for the ascription of blame and punishment.

Influence of distal event outcome on probability ratings.

Targeted analyses were examined whether participants saw the scenarios as structurally equivalent. The three conditional probability judgments were used to calculate the probabilistic contribution ratings for the distal (d) and proximal (p) events. These ratings reflect the degree to which each event raised the probability of the outcome occurring. For the distal event, this was the difference between the probability of the outcome (O), given the distal event did occur but the proximal event did not occur, where (n) denotes neither the distal or proximal event being present, $P(O|n+d)$, and the probability of the outcome prior to either the distal or proximal events occurred, [background conditions, $P(O|n)$]. Thus, the probabilistic contribution of the distal cause is $P(O|n+d) - P(O|n)$. For the proximal cause (p), this was the difference between the probability of the outcome (O), given both the distal and proximal events did occur [combined conditions, $P(O|n+d+p)$], and the probability of the outcome occurring given the distal event did occur but the proximal event did not occur, $P(O|n+d)$. Thus the probabilistic contribution of the proximal

cause is $P(O|n+d+p) - P(O|n+d)$. Analyses revealed no effects involving scenario, thus subsequent analyses were collapsed for scenario. Descriptive statistics are shown Table 3.4.

Table 3.4

Means and SDs for Probabilistic Contribution Ratings for Background, Distal, Proximal, and Combined Conditions by Condition (Study 3)

<i>Distal Event-Type Outcome</i>	<i>Background Conditions</i>	<i>Distal Event</i>	<i>Proximal Event</i>	<i>Combined Conditions</i>
Positive Action				
Positive Outcome	47.36 (29.14)	4.58 (43.86)	16.33 (42.02)	68.28 (29.62)
Negative Outcome	36.00 (26.68)	11.81 (39.08)	35.39 (32.87)	83.19 (18.33)
Negative Action				
Positive Outcome	58.78 (37.86)	-8.22 (54.53)	19.58 (46.72)	70.14 (32.94)
Negative Outcome	22.74 (23.48)	33.97 (29.09)	24.63 (24.76)	81.34 (15.43)
Physical Event				
Positive Outcome	54.72 (38.13)	-12.22 (50.11)	22.36 (46.24)	64.86 (31.75)
Negative Outcome	27.50 (27.50)	6.81 (33.75)	42.36 (32.87)	76.67 (21.01)

To test whether the effects reflect differences in the efficacy of the event to cause the outcome, an ANOVA was conducted with the between-subjects variables: 2 (outcome) and 3 (distal event) for the probabilistic contribution of the distal event. There was a main effect for outcome, $F(1, 209) = 15.33, p < .001, \eta^2 = .07$, but not for type of distal event. There was a 2-way interaction between outcome and distal event, $F(2, 214) = 3.10, p < .05, \eta^2 = .03$. This suggests that the effects do not reflect differences in participant's perception of the event's ability to cause the outcome, but do reflect the influence of the valence of the outcome on judgments of probability.

The analyses above shows that the valence of the outcome influences judgments, however to test whether the effects are diametrically opposed conditions, a MANOVA was conducted with two between-subjects variables: 2 (outcome) and 3 (distal event) for ratings of the probability of the outcome for the background conditions $[P(O|n)]$, and the combined conditions $[P(O|n+d+p)]$. The analyses revealed multivariate main effects for outcome, $V = 0.17, F(2, 208) = 20.46, p < .001, \eta^2 = .16$. Follow-up univariate analyses showed that the valence of the outcome influenced

probabilistic contributions for the background conditions, $F(1, 209) = 32.61, p < .001, \eta^2 = .14$, and the combined conditions, $F(1, 209) = 12.89, p < .001, \eta^2 = .06$. Participants' estimates of probabilistic contributions for the background conditions were *higher* for positive outcomes than negative outcomes, whereas the probabilistic contributions for the combined conditions were *higher* for negative outcomes than positive outcomes. There was no main effect for distal event, and no interaction between outcome and distal event.

These results show that probability judgments were influenced by outcome valence. The finding that the type of distal event did not influence judgments of probability for the background conditions, the distal event or after both events occurred, addresses concerns that participants perceived differences in the efficacy of each event to cause the outcome (Lombrozo, 2010).

Tests of Three Attribution Models.

Covariance analyses tested whether judgments of social controllability (Tetlock, 2002), praise/blame (Alicke, 2000), or the probabilistic contribution of the event (Spellman, 1997) explained causal judgments better than valence effects. The dependent variable was the extent to which the proximal event was preferred to the distal event (i.e., the difference between them), see Appendix E2 for descriptive statistics. The analyses statistically removed the dependent variable ratings from the interaction between outcome and distal event on causal ratings, $F(2, 209) = 8.89, p < .001$, partial $\eta^2 = .08$.

Tetlock's (2002) social functionalist model: Does punitiveness predict cause?

The results showed that judgments of social controllability, measured by ratings of punitiveness ($\beta = .04$) did not influence causal judgments.

Alicke's (2000) culpable control model: Does praise/blame predict cause?

The results showed that judgments of praise/blame ($\beta = .17$) did influence causal judgments, $F(1, 208) = 7.08, p < .01$, partial $\eta^2 = .03$. However, after covarying out participants' ratings of praise/blame on judgments of cause, the interaction between outcome and distal event remained significant, $F(2, 208) = 10.16, p < .001$, partial $\eta^2 = .09$. This suggests that people's moral judgments of another's action do influence their causal attributions; however the type of the distal event and the valence of the outcome have the greater effect.

Spellman's (1997) crediting causality model: Does probability increase predict cause?

The results showed that judgments of each event's probabilistic contribution ($\beta = .05$) did not influence causal judgments.

Together, these results show the utility in exercising social control through reward or retribution (Tetlock, 2000), people's desire to blame (Alicke, 2002), or the event that increases the likelihood of the outcome the most (Spellman, 1997) does not explain causal judgments in these data.

General Discussion

Studies 2 and 3 replicate the findings of Study 1 showing that the type of event (voluntary or physical), and the valence of the agent's motive and outcome have diverging effects on people's judgments of cause and blame. Importantly, the finding that people see intentional actions as more causal and blameworthy than physical events replicates earlier findings when the action is negatively motivated and harm eventuates (Lagnado & Channon, 2008; McClure et al., 2007). However, these studies add an important clarification: when the action is positively motivated or the outcome is positive, people's preference for intentional actions over physical events disappears. Instead, people rate both events as equally good causes when things end well, but discount positive actions in favour of physical events when things go awry.

The stability of the effects for outcome valence and type of distal event on judgments of cause and blame across Studies 1-3 addresses several questions. The bipolar praise/blame rating scale used in Study 2 and 3 addresses concerns that participants' responses may have been biased in Study 1 by a measure defined as 'blame' for ratings of physical events and actions in the positive conditions. The patterns of judgments of praise/blame from Study 2 and 3 replicate those reported in Study 1. This suggests that the new measure allowed participants to respond in a more nuanced manner to positively valenced conditions and physical events without undermining the robustness of the valence effects; negative actions were blamed, positive actions received at best a modicum of praise, and physical events were rated relatively neutrally unless things went wrong; in which case they were blamed.

The measure of responsibility was added in Studies 2 and 3 to investigate whether the effects in Study 1 reflect linguistic rather than psychological differences. The findings showed that judgments of cause, responsibility, and blame differ along important dimensions, but not exactly as predicted. People see cause and blame as distinct constructs when judging sequences comprising voluntary actions, but not when the sequence comprises only physical events. In this case, people do use the term 'blame' to describe a causal relation between events and effects in the absence of voluntary actions. Interestingly, the opposite pattern emerged for cause and

responsibility, where participants distinguished between physical events, but not voluntary actions.

The bipolar reward/punishment measure in Study 3 allowed for a stronger examination of the relations between cause and responsibility by testing how well several theories account for the findings in chains that include positive actions and outcomes. For example, attributional and legal theories claim that judgments of punishment reflect attributions of blame or responsibility to an agent: as judgments of responsibility and blame increase, so too do judgments of punishment (Hart & Honoré, 1985; Tetlock, 2000; Weiner, 1995). This idea suggests that the positive relations between these judgments should be mirrored in the positive conditions: as judgments of responsibility and praise increase, so too do judgments of reward. The findings in Study 3 supported these predictions and replicated the positive relation between responsibility and punitiveness found by Tetlock et al. (2007) for negative outcomes, but not for positive outcomes. Indeed, in chains with positive outcomes, judgments of cause and responsibility were strongly related, but neither construct related to judgments of praise, and only judgments of cause related to judgments of reward. These findings challenge several theoretical assumptions and suggest that the theorists (e.g., Tetlock, 2002; Weiner, 1995) were mistaken in their assumption that the relations between these constructs generalise to chains with positive outcomes.

The results from Studies 2-3 are consistent with research suggesting that people place different weightings on information about the morality of the action, the type of event (as shown by temporal order effects) and the outcome when judging cause, responsibility, blame and punitiveness (Alicke, 2000; Cushman, 2008; Hilton & McClure, 2013). Judgments of praise/blame reflected the morality of the action, whereas judgments of punitiveness relied jointly on the morality of the actions and the valence of the outcome. In comparison, judgments of cause and responsibility depended on all three factors, but the judgments differed in the relative importance or weighting of each factor: judgments of cause were influenced firstly by the valence of the outcome, and secondly, by the morality of the actions, whereas judgments of responsibility were influenced equally by the valence of the outcome, the voluntariness of the event, and the morality of the actions.

These findings make sense when we consider that judgments of cause reflect the relation between an event and an effect, whereas judgments of responsibility reflect a non-emotive assessment of accountability which does not presuppose judgments of liability for harmful outcomes. Furthermore, these findings show that people consider both the agent's intent and the

consequences of the events when making judgments of retribution. Again, this makes sense given the importance and implications of such judgments when people ascribe punishment or rewards in both courts of law and in the workplace. However, the finding that judgments of blame and punishment were not amplified for negative actions challenges the legal claim that the successful realisation of intended harm is twice as bad as intended harm that does not eventuate (Moore, 2009).

These results also demonstrate the pervasive influence of negativity by comparing ratings for voluntary actions against physical events in the baseline condition. In chains with negative outcomes, the valence of the action provokes responses that differ markedly from the baseline condition: negative motives amplify judgments of responsibility when harm eventuates even though people see the action as no more causal than physical events. In contrast, positive motives mitigate judgments of cause and people see the action as no more responsible than physical events. With positive outcomes, ratings of cause and responsibility for the distal physical event (the baseline condition) do not differ from ratings for positive or negative actions. This finding suggests that when things end well, people's responses are not influenced by the valence of the action. The finding that the valence of the outcome influenced judgments of praise/blame for physical events challenges the philosophical maxim that blame, being a moral evaluation, cannot be attributed to physical events (White, 1990).

The measure of probability added to Study 3 clarified first, whether the findings in Study 1 reflected non-equivalent outcomes and second, the ability of several theories to explain the findings. The results suggest that participants did not see positive outcomes (e.g., saving dozens of lives) or negative outcomes (e.g., losing dozens of lives) as diametrically opposing conditions. Indeed, people expect positive outcomes to be the most likely scenario until an event occurs, after which a negative outcome is the most likely occurrence. This reversal in reasoning suggests that peoples' evaluative reaction to harm rather than differences in the scenario paradigm *per se* is amplifying attributions.

The findings from Study 3 also suggest that neither probabilistic (e.g., Spellman, 1997) nor the social functionalist (Tetlock, 2002) approaches are able to explain why actions performed with negative intent are seen as more causal than the same actions performed with positive intent or physical events with the same causal effect. Indeed, in these causal chains, people did not see voluntary actions (e.g., a man lighting a fire) as a more reliable method of bringing about the outcome than physical events (e.g., a lightning strike). In contrast, replicating Studies 1 and 2,

when we compare ratings for voluntary actions leading to negative outcomes, the findings suggests Alicke (2000) was right to propose that people would amplify causal ratings for negative actions with negative outcomes more than positive outcomes. But these data challenge Alicke's claim that people's negative evaluative responses to a harmful outcome would have no effect on judgments for positive actions. Indeed, when we compare these ratings with the baseline condition of two physical events, people discount the causal contributions of positive actions to support attenuated blame attributions, but do not amplify the causal contribution of negative actions to support heightened blame attributions.

In sum, the findings from Study 2 and 3 go some way to addressing several questions regarding the structural equivalence of the conditions across the scenarios. Importantly, the findings also suggest that the effects of the valence of the agent's motive and outcome on judgments of cause and blame in Study 1 are not an artefact of the measures used, but that people see chains with positive outcomes very differently to those with negative outcomes.

Chapter Four: Explaining the Effects: Abnormality or the Facilitative Framing of Events?

The findings in Studies 2 and 3 show that attributions of cause, responsibility, and praise/blame are influenced by the type of event (voluntary or physical), the agent's motive, and the outcome of the events in these causal chains. These variables have different effects on different attributions: judgments about cause depended principally on the valence of the outcome, whereas judgments about praise/blame largely reflected the valence of the agent's motive. In contrast, judgments of punitiveness jointly reflected the valence of the agent's motive and the outcome, whereas judgments of responsibility were equally influenced by the voluntariness of the event, the valence of the agent's motive, and the outcome. Studies 1-3 showed that these effects are not artefacts of the measures used (Study 2), or differences in the efficacy of the events, or salience of the outcome (Study 3). Previous research has shown that the abnormality and the proximity of the events to the outcome also influence attributions (Gerstenberg & Lagnado, 2012; N'Gbala & Branscombe, 1995; Segura & McCloy, 2003; Wells et al., 1987). It is possible that the effects in Studies 1-3 reflect these factors rather than voluntariness or valence *per se*. Chapter 4 examined whether abnormality, as suggested by psychological and legal theorists (Hart & Honoré, 1985; Kahneman & Miller, 1986) can account for the effects of valence on people's attributions of cause.

Chapter 4 also addresses two questions about the exemplars of each category of event in these scenarios. The first question considers whether participants saw the physical events as good exemplars of that category of events. Replicating earlier research on causal chains (e.g., McClure et al., 2007), abnormality was implied in the exemplars of the physical event conditions in these studies (e.g. lightning strikes), and specific hypothesis were made based on this factor. For example, based on Hart and Honoré's (1985) claims about chains comprising two comparable events (e.g., or two actions or two physical events), McClure et al. predicted that people would prefer the most recent event. Yet, counter to Hart and Honore's claims, McClure et al. showed no difference between chains of two voluntary actions or two physical causes. Consistent with McClure al.'s findings, the results for the baseline condition with two physical events in these studies, show both events were rated equally causal although the proximal event (e.g., strong winds) was seen as increasing the outcome's probability more than the distal event (e.g., lightning strike). One explanation is that the effect of proximity was counterbalanced by the abnormality of the cause because people saw the distal event (e.g., lightning strike) as more

abnormal than the proximal event (e.g., strong wind). Thus the interaction between the events' abnormality and proximity may have influenced participants' weightings of voluntariness, motive, and outcome information. This question is investigated in Study 4.

The second question considers whether the structure of the causal chain, in particular the facilitative framing of the events, could explain the effects of valence on attributions of cause. Research suggests that people distinguish between events that facilitate the likelihood the outcome will occur, and those that inhibit the likelihood of the outcome occurring (Kelley, 1973; Mandel & Vartanian, 2009; Spellman et al., 2001). Research also demonstrates the impact of extra-evidential information, such as the observer's own norms, traits, biases, on attributions (e.g., Baron & Hershey, 1988; Enzle & Hawkins, 1992; Tetlock et al., 2010). Thus, it is possible that participants' evaluative reactions to the scenario lead them to amplify the facilitative features of the action rather than reflecting valence *per se*. For example, it would be reasonable to assume that a Conservation officer is likely to be regarded positively or at least neutrally, when their actions are aimed at inhibiting a negative outcome. In contrast, an arsonist is likely to be evaluated negatively, as the action was aimed at facilitating a harmful outcome. This question is addressed in Study 5. In addition, Chapter 4 also examined whether these effects generalise across different situations.

Pilot Studies.

Two further questions were investigated in two pilot studies (see Appendix F). Pilot Study 1 examined whether the effects in Studies 1-3 reflect differences in the type of motivation described in the action conditions. Research suggests that people's causal judgments reflect whether the action was motivated for instrumental reasons, such as personal gratification (e.g., revenge) or personal cost (e.g., risk life), rather than situational reasons, such as fulfilling employment contracts or acting in self-defence (e.g., Reeder, Kumar, Hesson-McInnis, & Trafimow, 2002; Reeder & Trafimow, 2005). Reeder et al. (2002) showed that people rated the same aggressive action (e.g., a punch) more causal and blameworthy when it was motivated for instrumental (e.g., reprisal) rather than situational (e.g., self-defence) reasons. Thus, participants in Studies 1-3 might have given higher causal judgments because they inferred that the action was motivated by instrumental rather than situational reasons. For example, participants might have considered the arsonist's action as motivated for instrumental reasons, whereas the Conservation officer's action could reasonably be perceived as part of their normal job description thus situational motivated.

Pilot Study 1 presented participants with a scenario describing a distal action motivated for instrumental or situational reasons, and a proximal physical event leading to positive or negative outcomes. The action described the same behaviour that was either positively or negatively motivated. Each scenario was followed by a series of questions that elicited judgments of goodness/badness and cause. The results showed that participants saw the same intentional action as bad when performed with negative intent, but good when performed with positive intent. Participants' judgments of cause however, were not sensitive to the type of motivation.

Pilot Study 2 investigated whether differences in the type of physical events used as exemplars could explain the effects in Studies 1-3. Specifically, did participants see mechanical events (e.g., faulty wiring) as more causal than natural events (e.g., wind or lightning strike) based on perceptions of greater controllability and preventability? The effect of controllability on attributions is well established; people judge controllable factors, such as an agent's decision (Kahneman & Miller, 1986), as more causal than uncontrollable factors (Giroto, Legrenzi, & Rizzo, 1991). Research on the effects of morality and norm violations on attributions has shown that people's judgments of cause and blame reflected the extent that the event violated functional norms rather than voluntariness *per se* (Creyer & Gürhan, 1997; Hitchcock & Knobe, 2009; Morris, Moore, & Sim, 1999). For example, Hitchcock and Knobe (2009) found that participants rated events caused by mechanical failures to be as blameworthy as those caused by human error. This implies that events caused by mechanical failure reflect human error, so are seen as more controllable than natural events (Naquin & Kurtzberg, 2004). Therefore, it is possible that judgments about the events in these causal chains were influenced by norm violations (moral, functional, or statistical), rather than differences between voluntary actions and physical events.

Pilot Study 2 presented participants with a scenario describing two consecutive events leading to a positive or negative outcome. Both events had two conditions; a mechanical event, or a natural event that had equivalent effects on the outcome. Each scenario was followed by a series of questions that elicited judgments of cause, abnormality, and goodness of explanation. The results showed participants did not see mechanical events as better explanations than natural events for the outcomes.

In sum, these findings suggest that people do see positive and negative actions differently, but that the effects reported in earlier studies were probably not due to differences in the type of motivation driving the action, or the physical event exemplar used in the scenarios.

Study 4: Does Abnormality Explain Attributions?

The central focus of Study 4 was twofold. First, to examine whether the abnormality of the events can explain the causal relations between successive events reported in Studies 1-3. To investigate this possibility, the physical event category was extended to include normal and abnormal conditions, and all event exemplars were revised to meet the legal criteria¹¹ for each type of event (Hart & Honoré, 1985; Moore, 2009). A measure of abnormality was introduced to examine lay attributions in chains when voluntary actions and physical events are pitted against each other. Based on Hart and Honoré's (1985) claim about causal chains that include voluntary actions and independent physical events, I predicted a 2-way interaction between the type of event and temporal position for judgments of cause. Specifically, regardless of the outcome, voluntary actions would be rated *higher* than normal proximal physical events but *lower* than abnormal proximal physical events. Second, the most recent abnormal physical event would be rated higher than normal physical events or preceding abnormal events. The second aim was to investigate whether the effects reported in Studies 2-3 would generalise across differing contexts. Study 4 therefore examined these effects adding two new scenarios.

Method

Design.

Two changes were made to the design used in Studies 1-3. First, two new scenarios (avalanche and flooding) were added in Study 4, making a total of five scenarios (See Appendix A, pp. 133-142). Second, the distal and proximal physical events had two versions: a normal or an abnormal event. The physical event was the same event in both normative conditions. Thus, Study 4 extended the mixed design used in Studies 1-3, to include four between-subject variables: distal event (positive action, negative action, normal physical, or abnormal physical), proximal physical event (normal or abnormal), and outcome (positive or negative), and scenario (fire, train, factory, avalanche, flooding). A fifth within-subject variable was the temporal position of the event (distal or proximal).

Scenarios.

Two changes were made to the physical event exemplars used in Study 4. First, to satisfy the legal criterion for abnormal physical events to be 'extraordinary for the time and place', abnormal physical events were described as a 'rare occurrence', whereas normal physical events

¹¹ See Chapter 1, p. 4 for a discussion (Hart & Honoré, 1985).

were described as a ‘frequent’ occurrence. Thus, the abnormality of the distal physical event was manipulated using the word ‘*frequently*’ in the normal condition or ‘*rarely*’ in the abnormal condition. The proximal physical event was same type of event in all conditions (e.g., wind); its normality or abnormality however was manipulated by the event’s description. For example, in the fire scenario, the normal version of the proximal physical event was ‘*a strong wind*’, whereas the abnormal version referred to the wind as a ‘*tornado*’. Second, for consistency, all physical events were revised to describe natural rather than mechanical events.

Two changes were made to the structure of the scenarios in Study 4. In Study 3, the measure of probability showed that participants saw positive outcomes as moderately probable, whereas negative outcomes were highly improbable before either event in the chain occurred, yet the reverse was true after both events had occurred. It is unclear if these effects reflect differences in the experimental manipulations or whether participants see positive and negative outcomes differently. To minimise the possible effects of negative outcomes being over-weighted in attributional decisions, two changes were made to emphasize the extent that the realised outcome deviated from the normal or expected state of affairs due to the sequence of events.

First, each scenario was prefaced with a background statement which aimed to minimise any outcome bias by stating the normative condition prior to the distal and proximal events occurring (Alicke et al., 2011). In the fire scenario, for example, the background statement read: *A forestry town was threatened by wildfires on several fronts*. Second, the outcomes were revised to describe how the positive or negative outcome diverged from the normative condition detailed in the background statement. In Studies 1-3 for example, the positive outcome in the fire scenario read: *These events prevented the wildfire reaching the nearby town, saving dozens of homes*. In Study 4 the revised statement in the same condition read: *These events prevented the wildfire from reaching the nearby town, saving the lives of dozens of people trapped in their homes*.

Measures and procedure.

Participants ($N = 230$) read a single scenario then gave judgments of the distal and proximal events’ of cause, responsibility, praise/blame, punitiveness, and probability on a rating scale. The measures of cause, responsibility, praise/blame, punitiveness, and probability were the same as Study 3. The measure of abnormality replicated that used in the pilot studies and was based on (Ryder, 1997). The measure read: ‘Please rate the extent to which you think each of the following events is unusual’. The stem was followed by rating scales for each event, where 0 = Very unusual and 10 = Not at all unusual.

The measures of cause, responsibility, praise/blame, punitiveness and abnormality were counterbalanced in a Latin square design. Participants completed the probability task on a separate page. The order of probability attributions was reversed for half of the conditions. The order of the dependent variables did not influence ratings, so subsequent analyses were collapsed for this factor.

Results

Effects of outcome valence and types of events across multiple scenarios.

To test whether the findings in Studies 1-3 were replicated with more scenarios, two mixed design ANOVAs were performed for judgments of cause and estimates of the probability of the outcome. In both analyses, the four between-subjects variables were: 5 (scenario; fire, train, factory, avalanche, flooding) x 2 (outcome: positive, negative) x 4 (distal event: positive action, negative action, normal physical, abnormal physical), and 2 (proximal physical event: normal, abnormal). The first ANOVA was performed with the within-subjects variable: 2 (temporal position: distal, proximal) for judgments of cause as the dependent variable. The second ANOVA was performed with the within-subjects variable: 4 (temporal position: background conditions, distal event, proximal event, combined conditions) with judgments of outcome probability as the dependent variable. There was no 5-way interaction between scenario and these variables for either analysis, so subsequent analyses were collapsed for scenario.

Judgments of abnormality.

As a manipulation check on whether participants saw abnormal physical events (e.g., tornados) as more abnormal than normal physical events (e.g., strong winds), a MANOVA was performed with three between-subjects variables: 2 (outcome), 4 (distal event), and 2 (proximal event) for ratings of the distal and proximal events for judgments of abnormality (see Figure 4.1 and Appendix G1). A main effect for distal event, showed that negative actions were rated more abnormal than abnormal physical events, which in turn were rated more abnormal than positive actions and normal physical events, which did not differ, $F(6, 1782) = 23.94, p < .001, \eta^2 = .08$. There was also a main effect for proximal event, where abnormal physical events were rated higher than normal physical events, $F(2, 890) = 5.32, p < .02, \eta^2 = .02$. There was no main effect for outcome or interactions.

Figure 4.1. Means for abnormality ratings for distal and proximal events by condition. Error bars show standard errors (Study 4).

Influence of distal event, abnormality, proximity, and outcome on attributions.

To examine whether the valence effects reported in Studies 1-3 were replicated in Study 4, a MANOVA was performed with three between-subjects variables: 2 (outcome), 4 (distal event), and 2 (proximal event) for the repeated measures of cause and praise/blame. See Figure 4.2 and Appendix G1 for descriptive statistics. The analyses revealed multivariate main effects for outcome, $F(4, 888) = 57.46, p < .001, \eta^2 = .21$, and distal event, $F(12, 2670) = 28.76, p < .001, \eta^2 = .11$. The analyses also revealed two multivariate 2-way interactions between outcome and distal event, $F(12, 2670) = 4.70, p < .001, \eta^2 = .02$, and proximal event, $F(12, 2670) = 2.00, p < .03, \eta^2 = .01$, but there was no 3-way interaction between outcome, distal event and proximal event. Replicating Study 3, the main effects were qualified by the same 2-way interactions reported above for judgments of cause, but not praise/blame. Follow-up analyses showed the same pattern and effect sizes for judgments of cause and praise/blame closely replicated Studies 1-3, so are not discussed further.

The relations between cause, responsibility, praise/blame and punitiveness.

To examine whether the relations between judgments reported in Studies 2-3 were replicated in Study 4, correlational analyses used the difference score between ratings of the distal and proximal events for positive and negative outcomes (see Table 4.1). Replicating Study 2 and 3, participants saw judgments of cause and responsibility, and judgments of praise/blame and punitiveness as two discrete groups of constructs regardless of the outcome. Therefore, subsequent analyses focus on judgments of cause and praise/blame.

Figure 4.2. Mean ratings of cause, responsibility, praise/blame and punitiveness for distal and proximal events by condition (Study 4).

Table 4.1

Means, SDs and Correlations for the Difference Between the Proximal and Distal Events Cause, Praise/Blame, Responsibility, Punitiveness and Probability ratings by condition (Study 4)

	<i>M (SDs)</i>		Cause		Responsibility		Praise/blame	
	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome
<i>Cause</i>	12.76 (37.86)	8.85 (34.69)	-	-				
<i>Responsibility</i>	6.73 (39.36)	2.64 (37.05)	.62	.73	-	-		
<i>Praise/Blame</i>	-13.02 (26.44)	-3.52 (25.84)	-.18	.61	-.18	.60	-	-
<i>Punitiveness</i>	-9.54 (25.41)	-8.90 (20.99)	-.19	.42	-.19	.45	.64	.67

All correlations were significant, $p < .001$.

Test of Psychological Theory: Does Abnormality Predict Cause?

To test whether the abnormality of the event can account for the key effects, a between-subjects ‘type of chain’ variable was created by crossing the four distal event and two proximal event conditions. Thus, the type of chain variable had eight conditions that described the distal and proximal events. Four conditions comprised voluntary actions: positive or negative actions followed by normal or abnormal physical events. Four comprised two physical events: normal or abnormal distal physical events followed by normal or abnormal proximal physical events. The dependent variable was the extent to which the proximal event was preferred to the distal event (i.e., the difference scores), see Appendix G2 for descriptive statistics.

Covariance analysis was conducted with two between-subjects variables: 2 (outcome) x 8 (chain) to test whether the abnormality of the event better explains causal judgments than valence *per se* (Kahneman & Miller, 1986). The analysis removed abnormality ratings from the main effect for chain, $F(7, 891) = 3.38, p < .01$, and the 2-way interaction between outcome and chain, $F(7, 891) = 8.25, p < .001$, for judgments of cause. The results showed that judgments of abnormality ($\beta = .24$) had a strong effect on causal judgments, $F(1, 890) = 47.58, p < .001$. After covarying out ratings of abnormality on judgments of cause, however, the main effect for chain-type, $F(7, 890) = 2.52, p < .02$, and the interaction between outcome and chain remained significant, $F(7, 890) = 8.19, p < .001$. These results show that judgments of the abnormality of an event do influence causal attributions, but voluntariness, the agent’s motive, and the valence of the outcome remain the most influential factors when people ascribe cause.

Test of Legal Theory: Does Abnormality and Proximity Predict Cause?

To test Hart and Honoré’s (1985) claim that abnormal proximal physical events are rated higher than distal intentional actions, a mixed design ANOVA was performed with two between-subjects variables: 2 (outcome) x 8 (chain) and one within-subjects variable: 2 (temporal position) for judgments cause. Counter to the predicted 2-way interaction between chain and temporal position, there was a 3-way interaction between outcome, chain and temporal position, $F(7, 891) = 8.25, p = .001, \eta^2 = .06$. See Appendix H for full analyses. These results do not support Hart and Honoré’s claim that when chains comprise voluntary actions and physical events, people prefer the most recent voluntary action, unless that action is followed by an independent abnormal physical event. These results are also counter to their claim that when chains comprise two equivalent types of events (as in the baseline condition), people will prefer the most recent event.

The influence of outcome valence on probability ratings.

Targeted analyses of the baseline condition of chains with two physical events tested whether the effects of outcome valence on probability judgments reflect a bias towards negative outcomes. See Table 4.2 for descriptive and inferential statistics. T-tests showed participants anticipated positive rather than negative outcomes prior to either event occurring in all conditions. In contrast, after both events occurred, participants saw negative outcomes as likely as positive outcomes when the normative conditions in the chain were the same, or ‘matched’ (e.g., Normal:Normal; Abnormal:Abnormal), but they saw negative outcomes as more likely when the normative conditions were incongruent, or ‘mismatched’ (e.g., Normal:Abnormal; Abnormal:Normal). The finding that participants rated positive outcomes more likely than negative outcomes prior to either event occurring suggests that people expect things to turn out well, whereas, the finding that they saw either outcome as equally likely after both physical events occurred suggests that their judgments were not biased by knowing negative outcome information. This finding addresses concerns that participants’ responses were biased by the valence of the outcome.

Table 4.2

Means and SDs for Ratings of the Probability of the Outcome Occurring by Condition (Study 4).

Variable: Chain-Type	Background Conditions			Combined Conditions		
	Positive Outcome	Negative Outcome	<i>t</i>	Positive Outcome	Negative Outcome	<i>t</i>
Distal event: Proximal event						
Normal Physical Event:						
Normal event	57.22 (33.67)	24.64 (29.85)	5.38***	71.30 (25.03)	80.00 (23.12)	1.90
Abnormal event	52.68 (34.08)	16.32 (23.04)	6.60***	65.18 (26.08)	79.30 (23.67)	3.02**
Abnormal Physical Event:						
Normal event	59.64 (35.48)	15.34 (14.29)	9.00***	68.70 (30.03)	80.69 (20.25)	2.49*
Abnormal event	59.64 (35.48)	23.27 (25.68)	6.16***	73.82 (25.64)	71.82 (28.75)	0.39

* $p < .05$. ** $p < .01$. *** $p < .001$. $df = 108 - 114$.

Discussion

The results in Study 4 provide further support for the robustness of the effects in earlier studies and extend the findings to two new scenarios. As in Studies 1-3, judgments about cause

reflected voluntariness, motives and outcomes, whereas judgments about praise/blame primarily reflect motives.

Study 4 examined whether the abnormality of events, as proposed by legal theorists (Hart & Honoré, 1985; Moore, 2009), might explain the effects of valence on lay judgments of cause. The results suggest that people do apply the legal criterion that a physical event must be extraordinary for both the time and place to be considered abnormal (Hart & Honoré, 1985; Moore, 2009). Interestingly, the results also show that people see negative actions as abnormal events, yet saw positive actions as normal events even when those actions are performed knowing a high personal cost would be realised. In the flood scenario for example, the engineer knew his home would be destroyed if he blew up that section of levee. This asymmetry in judgments for positive and negative conditions is consistent with valence research showing that people expect people's action to conform to social norms (Ybarra, 2002).

Importantly, these findings show that the abnormality of the events does influence people's attributions in these causal chains (Kahneman & Miller, 1986), but abnormality was not a core determinant when ascribing cause. Instead, replicating Studies 1-3, their causal ratings reflected the extent that the valence of the agent's motive matched that of the outcome. Further, the abnormality of the event did not influence judgments of the likelihood that the outcome would occur. Indeed, regardless of how the chain ended, participants saw proximal physical events as contributing more to the likelihood that the outcome would occur than distal events.

These findings challenge the legal descriptions of the conditions necessary for an abnormal physical event to 'break' the causal connection between voluntary actions and the outcome (Moore, 2009). Specifically, participants did not consistently see voluntary actions as more causal than normal physical events, or abnormal physical events that occurred later in the chain as more causal than preceding voluntary actions. For example, in chains comprising voluntary actions and physical events, contrary to the predicted 2-way interaction between voluntary actions and proximal physical events, there was a 3-way interaction between outcome, voluntary actions and proximal physical events. When things ended well, participants saw normal as well as abnormal physical events that occurred later in the causal chain as the primary cause of good outcomes, regardless of the preceding event. In legal terms, this suggests that participants saw both abnormal and normal physical events as intervening events that 'broke' the causal connection between voluntary actions and the outcome. However, when harm eventuated, participants' perception of the efficacy of subsequent physical events was trumped by the valence of the

agent's motive. Participants saw both normal and abnormal physical events as 'intervening events' when preceded by actions performed with positive intent. But abnormal proximal physical events were relegated to the background when the same action was performed with negative intent. This suggests that both abnormality and proximity play a less influential role than valence in lay judgments of cause, even when the events are equivalent in probabilistic terms.

Study 5: The Effect of the Facilitative Nature of the Event.

The results from Study 4 closely replicate those reported in Studies 1-3 while discounting several alternative explanations for the effects of motive and outcome valence on attributions. The findings suggest that neither the abnormality of the events (Hart & Honoré, 1985), nor the social functionalist model (e.g., Tetlock, 2002) or the probabilistic approaches (e.g., Spellman, 1997) are able to explain why actions performed with negative intent are seen as more causal than physical events or the same action performed with positive intent when harm occurs, but less causal when good eventuates.

However, it is possible that the facilitative framing of the events, rather than the core components of voluntariness, intentionality, or abnormality may explain these findings. Spellman et al. (2001) suggest people distinguish between events that heighten, or facilitate the likelihood the outcome will occur and those that decrease, or inhibit the likelihood of the outcome occurring. Recently, Mandel and Vartanian (2009) found that people weigh causal information depending on their assessment of whether the cause generated or prevented the outcome. Indeed, Kelley (1973) proposed that people's causal attributions are often informed by their schemas regarding the opposing effects of generative or preventative causes. Specifically, Kelley suggested that the occurrence of an action in the presence of a strong inhibitory cause is likely to lead to people augmenting the generative characteristics of the action.

It is possible therefore that the outcome statement in these studies may have influenced responses by describing the chain of events either '*lead to*' negative outcomes, or '*prevented*' positive outcomes. Alternatively, it is possible that judgments of cause and blame were influenced by the type of event because of differences in the generative or preventative effect of that event on subsequent events or the outcome. To address these questions, two revisions were made to the scenarios in Study 5. First, the outcome statements were revised to replace the terms '*lead to*' or '*prevented*' with the same neutral link between events and outcomes, e.g., '*These events resulted in the...*'. Second, the proximal physical event was revised to ensure the event made an independent, albeit ambiguous contribution to the outcome which did not inhibit or

facilitate the preceding event's effect. In the fire scenario, for example, the proximal event was changed from: '*That night a tornado developed which fanned the flames*' to read: '*That night a tornado struck, fanning the flames and uprooting several trees, diverting fire engines around blocked roads.*'

A second related issue regards the voluntary action descriptors. Research has shown that the characteristics of the agent influence people's judgments, and also that the observer's own biases and ideologies influence their attributions (Baron & Hershey, 1988; Hawkins & Hastie, 1990; Tetlock et al., 2007). Indeed, both Alicke's (2000) Culpable Control model and Tetlock's social functionalist model (2002) emphasize the evaluative impact that extra-evidential information, such as the observer's own values, traits, biases, etc., can have on attributions.

Perhaps, then, the reported effects of the agent's motive in judgments of cause and blame in Studies 1-4 were shaped by the participants' own evaluative reactions and world knowledge rather than the valence of the agent's motive or the outcome *per se*. For example, it would be reasonable to assume that a Conservation officer is likely to be regarded positively or at least neutrally, whereas an arsonist is likely to be evaluated negatively and considered counter-normative. To address this concern, the scenarios were revised to describe the agent neutrally and in a role that does not violate professional (e.g., *an employee, a man, or an engineer*) or legal norms (e.g., arson, illegal drug activities, or traffic infringements).

The measures of punitiveness and responsibility were omitted from Study 5 for two reasons. First, the main effects and interaction terms were almost identical across Studies 2- 4. Second, there were strong positive correlations ($p < .001$) between the measures of cause and responsibility (between-studies median $r = .59$, mode $r = .62$), and between blame and punitiveness (between-studies median $r = .50$, mode $r = .59$) but not between cause and blame, or responsibility and punitiveness in Studies 2- 4. Taken together, these findings suggest that although distinct, the differences between these constructs are relatively subtle.

Method

Design.

Two changes were made to the design used in Study 4. First, three more new scenarios (soldier, casino, and car) were added in Study 5, see Appendix A, pp. 143-147). The soldier scenario adapted Alicke et al.'s (2010) extension of Hitchcock and Knobe's (2009) physician scenario. The scenario was re-written to minimise ratings being influenced by perceived violations of professional standards (e.g., the Hippocratic Oath) rather than valence *per se*. The

car scenario was adapted from Alicke (1992, Expt. 1). Although scenario did not interact with any factors in Study 4, the flooding scenario did not function well so was omitted.

Second, Study 4 showed that although people do distinguish between abnormal and normal physical events, the findings also showed that attributions reflect the structure of the chain and outcome rather than the abnormality of the physical event. To ensure equivalence across the conditions while reducing complexity in the study's design, the normal physical event in Study 4 was omitted in Study 5. Thus, the distal event had three conditions that had equivalent effects on the outcome; a positive action, a negative action, or an abnormal physical event. The proximal event was an abnormal physical event that was the same in all conditions.

Thus, as in Studies 1-3, Study 5 used a mixed design with three between-subject variables: type of distal event (positive action, negative action, or abnormal physical event), outcome (positive or negative) and scenario (fire, train, factory, avalanche, car, casino, or soldier), and a fourth within-subject variable: the temporal position of the event (distal or proximal).

Scenarios.

Three changes were made to the scenarios in Study 5. First, the outcomes were revised to include a neutral link between events and outcomes. Second, the proximal physical event was revised to ensure that it made an independent contribution to the outcome. Third, the scenarios were rewritten to describe the agent in the same neutral role. In the fire scenario, for example, the distal action statement was revised to read: *A man deliberately set fire to an area of forest* rather than *An arsonist deliberately set fire to an area of forest*.

Measures and procedure.

Participants ($N = 357$) were randomly assigned to either positive or negative outcome conditions and received four different scenarios. Each scenario was followed by a series of questions that elicited judgments of cause, praise/blame, and probability. The measures of cause, praise/blame, and probability were the same as Study 4 and were counterbalanced in a Latin square design. Analyses revealed no effects for order, or 3-way interaction between scenario, distal event and outcome for judgments of cause or probability, so subsequent analyses were collapsed for order and scenario.

Results

A MANOVA was conducted with two between-subjects variables: 2 (outcome: positive, negative), 3 (distal event: positive action, negative action, physical event) for the repeated measures of cause and praise/blame. See Figure 4.3 and Appendix I1 for descriptive statistics.

The analyses revealed multivariate main effects for outcome, $F(4,348) = 28.22, p < .001, \eta^2 = .25$, and distal event, $F(8, 698) = 13.05, p < .001, \eta^2 = .13$. There was a 2-way interaction between outcome and distal event, $F(8, 698) = 3.29, p < .001, \eta^2 = .04$. Follow-up analyses (see Table 4.3) showed the same pattern for judgments of cause and praise/blame as in Studies 2-4, so are not discussed further.

Table 4.3.

Summary of Inferential Statistics for Judgments of Cause and Praise/Blame (Study 5).

Variable <i>Temporal Position</i>	Outcome		Distal event		Outcome x Distal event	
	<i>F</i> (1, 891)	<i>p</i>	<i>F</i> (2, 351)	<i>p</i>	<i>F</i> (2, 351)	<i>p</i>
Cause						
<i>Distal</i>	0.01	<i>ns</i>	2.00	<i>ns</i>	5.31	.010
<i>Proximal</i>	0.64	<i>ns</i>	2.03	<i>ns</i>	6.63	.001
Praise/Blame						
<i>Distal</i>	42.96	.001	37.87	.001	0.26	<i>ns</i>
<i>Proximal</i>	87.05	.001	5.88	.010	0.01	<i>ns</i>

Figure 4.3. Mean ratings of cause and praise/blame for distal and proximal events by condition (Study 5).

Influence of distal event, temporal position, and outcome on probability ratings.

A mixed design ANOVA examined whether the effects in Studies 1-4 were replicated when the facilitative framing of the events were equalised across conditions. The between-subjects variables were: 2 (outcome), x 3 (distal event), and the within-subjects variable was: 3 (temporal position: background, distal, proximal) for judgments of the probabilistic contribution of events to the outcome. See Figure 4.4 and Appendix I2 for descriptive statistics.

There was a main effect for outcome, $F(1, 351) = 14.31, p < .001, \eta^2 = .04$, temporal position, $F(2, 350) = 59.67, p < .001, \eta^2 = .25$, but importantly, there was no main effect for distal event, $F(2, 351) = 2.37, ns$. This finding suggest that when all else is held constant, participants did not see actions as facilitating or inhibiting the outcome more than physical events as measured by probability increase. Instead, participants' judgments reflected the positive or negative valence of the outcome and the agent's motive.

Figure 4.4. Means for probabilistic contribution ratings for background, distal, and proximal events by condition. Error bars show standard errors (Study 5).

There was also a 3-way interaction between outcome, distal event and temporal position, $F(4, 702) = 9.18, p < .001, \eta^2 = .05$. To clarify this interaction, and to investigate how judgments of each event's contribution to the outcome differed depending on the type of distal event, three mixed ANOVAs were conducted for each distal event. In the baseline condition with two physical events, the analyses revealed a main effect for temporal position, $F(2, 116) = 28.68, p < .001, \eta^2 = .33$, but not for outcome. Thus, participants did not see a positive or negative outcome as the more likely outcome, but how much each event contributed to the outcome differed

depending upon what eventuated. Participants saw background conditions as contributing *more* than either distal or proximal physical events regardless of outcome; however the contribution of distal and proximal physical events differed depending on the valence of the outcome. With negative outcomes, participants saw distal and proximal events as contributing *equally* and *increasing* the likelihood of the outcome. Yet, with positive outcomes, distal events *decreased* the likelihood of the outcome whereas proximal events *increased* the likelihood of the outcome.

With chains with positive or negative actions, there were main effects for outcome (positive actions, $F(1, 115) = 5.55, p < .03, \eta^2 = .05$; negative actions, $F(1, 119) = 10.23, p < .01, \eta^2 = .08$), and temporal position (positive actions, $F(2, 116) = 28.68, p < .001, \eta^2 = .33$; negative actions, $F(2, 118) = 20.06, p < .001, \eta^2 = .25$). These main effects were qualified by a 2-way interaction between outcome and temporal position for chains comprising positive actions, $F(2, 114) = 6.70, p < .01, \eta^2 = .11$, and negative actions, $F(2, 118) = 9.44, p < .001, \eta^2 = .14$.

As with ratings of events in the baseline condition, the same pattern of probabilistic increase was revealed; however, the pattern reflected the match between the agent's motive and outcome, rather than the valence of outcome alone. When the motive and outcome matched (e.g., positive motive: positive outcome, or negative motive: negative outcome), the pattern was the same as the baseline condition with negative outcomes; participants saw distal and proximal events as contributing *equally* and *increasing* the likelihood of the outcome. In contrast, when the motive and outcome were mismatched (e.g. positive motive: negative outcome, or negative motive: positive outcome), the pattern was the same as the baseline condition with positive outcomes; distal events *decreased* the likelihood of the outcome, whereas proximal events *increased* the likelihood of the outcome.

Tests of Two Attribution Models.

Given the increasingly robust designs across Studies 1-5, two sets of covariance analyses replicated those conducted in Study 3 to test whether judgments of praise/blame (Alicke, 2000), or the probabilistic contribution of the event (Spellman, 1997) explained causal judgments better than valence effects. The dependent variable was the extent to which the proximal event was preferred to the distal event (i.e., the difference between them). See Table 4.4 for descriptive statistics. The analyses removed ratings of praise/blame, or probabilistic contribution from the main effect for distal event, $F(2, 351) = 4.25, p < .02$, and interaction between outcome and distal event, $F(2, 351) = 12.32, p < .001$, for judgments of cause.

Table 4.4.

Means and SDs for the Difference between the Distal and Proximal Events for Ratings of Cause, Praise/Blame, and Probabilistic Contribution by Condition (Study 5).

Variable	Cause		Praise /Blame		Probabilistic Contribution	
	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome
Positive Action	3.39 (35.39)	27.42 (42.62)	7.81 (22.16)	15.58 (36.50)	-3.60 (51.31)	37.37 (69.47)
Negative Action	21.67 (41.27)	-1.12 (38.47)	-23.41 (29.73)	-17.59 (27.43)	47.00 (73.11)	8.88 (49.10)
Physical Event	-1.32 (29.71)	4.57 (30.32)	-1.56 (23.32)	1.90 (16.30)	29.20 (52.68)	8.05 (43.66)

Alicke's (2000) culpable control model: Does praise/blame predict cause?

Replicating Study 3, the analyses showed that judgments of praise/blame ($\beta = .31$) had a strong influence on causal judgments, $F(1, 350) = 31.92, p < .001$, partial $\eta^2 = .08$. However, after covarying out participants' ratings of praise/blame on judgments of cause, the main effect for distal event, $F(2, 350) = 6.39, p < .01$, and interaction between outcome and distal event, $F(2, 350) = 13.11, p < .001$, remained significant. This suggests that people's moral judgments of another's action do influence their causal attributions; however the type of the distal event and the valence of the outcome have the greater effect.

Spellman's (1997) crediting causality model: Does probability increase predict cause?

Unlike Study 3, the analyses showed that judgments of probabilistic contribution ($\beta = .24$) also had a strong effect on causal judgments, $F(1, 350) = 12.68, p < .001$. However, after covarying out ratings of probabilistic contribution on judgments of cause, the main effect for distal event, $F(2, 350) = 4.43, p < .02$, and the interaction between outcome and distal event, $F(2, 350) = 8.01, p < .001$, remained significant. These results show that judgments of each event's probabilistic contribution to the outcome does influence causal attributions, however, the valence of the agent's motive and the outcome remain the most important factors when people ascribe cause.

Discussion

Study 5 investigated whether the effects reported in Studies 1-4 might have been due to differences in the facilitative framing of the events. These results provide strong evidence against

this alternative account. For example, actions that could be described as generative causes, e.g., a man deliberately lighting a fire, were indeed seen as increasing or facilitating the intended outcome, but contrary to predictions, the same actions were seen as preventative causes that decreased or inhibited the outcome when things went awry. This finding suggests that Kelley (1973) and Spellman et al. (2001) were right to suggest that people distinguish between events that inhibit or facilitate the outcome, but this distinction does not explain their causal judgments or the effects of valence in these studies.

As in Study 3, participants' moral judgments, or the probabilistic contribution of the background conditions and each event to the outcome cannot explain why people see voluntary actions as the primary cause of bad outcomes but not good outcomes. Replicating earlier research on chains leading to negative outcomes (Hilton et al., 2010), voluntary actions are not seen as better causes than physical events by virtue of their effect on the probability of the outcome. In fact, actions were seen as no more efficacious than physical events in producing the outcome when they were performed with negative intent, yet they were seen as being significantly better causes than physical events when performed with positive intent.

These findings suggest that people can act as 'intuitive statisticians' and assess the extent each event raised (or decreased) the outcome's probability in these causal chains. But counter to predictions, the event perceived to most raise the probability of the outcome, was not selected as the primary cause of the outcome. Indeed, these results do not support Spellman's (1997) proposal that causal selection is comparable to multiple regression analyses in which each of the predictors (i.e., events) are entered into the model in chronological order (Hilton, 1988). Instead, the patterns suggest that people's judgments of probability, like judgments of cause, responsibility and blame, are shaped by the valence of the agent's motive and the outcome.

General Discussion

The distinct effects of outcome, type of event, and motivation information (as shown by the proportion of variability explained by each factor) on judgments of cause and praise/blame found in Study 4 and 5 closely replicate those found in Studies 1-3; judgments about cause depended principally on the valence of the outcome, whereas judgments about praise/blame largely reflected the valence of the agent's motive. Consistent with Cushman's (2008) work on the properties of moral judgments (e.g., permissibility, wrongness, punishment and blame), these findings suggest that judgments of cause, responsibility and praise/blame can also be

characterised in terms of a stable set of effects that generalise across several differing contexts and everyday scenarios.

These studies provide important evidence against several possible alternative explanations for the effects reported in Studies 1-3, using more robust designs. Two pilot studies conducted showed that the type of motive (e.g., instrumental or situational), or the physical event used as an exemplar of that category (e.g., a mechanical event or a natural event) did not negate the effects of positive or negative valence on judgments. Consistent with earlier research (Kahneman & Miller, 1986), Study 4 showed that the abnormality of events does effect causal attributions. However, these results demonstrate that while participants consider abnormality to be an influential factor, it is less important than voluntariness, motive, or outcome valence information in their causal attributions. Study 5 replicated the findings from Study 4, and showed that the effects of valence cannot be explained by the facilitative nature of the events.

The revisions made to the scenarios in Study 4 and 5 address the possibility that participants' responses were biased by the valence of the outcome, or participants' own biases. In both studies, the valence of the outcome had no effect on judgments of cause in the baseline condition of two physical events. This finding suggests that when all else is equal, participants did see positive and negative conditions as diametrically opposed. In Study 5, the type of event had no influence on judgments of cause or probabilistic contribution of distal events. This finding suggests that when valence was controlled, participants did see physical events as having the same causal effect to equivalently located voluntary actions. Furthermore, these findings show that any divergence from the response patterns reported in the baseline conditions of two physical events demonstrates the effects of voluntariness and the valence of the agent's motive and outcome on people's judgments of cause.

Taken together, the findings from Study 3, 4, and 5, suggest there are two tiers of influential factors in people's attributional decisions. At the first tier, the factors of voluntariness, motive and outcome information appear to be the core determinants. At the second tier, the factors of proximity, abnormality, and probabilistic contribution, appear to play a secondary, albeit influential role in lay attributions. Importantly however, these findings add further evidence to the notion that it is the extent of 'match' between the agent's intent and what eventuated that shapes judgments of cause, whereas it is the agent's intent that determines judgments of culpability.

Chapter Five: The Role of Intentionality and Foresight on Attributions

Studies 1-5 tested the predictions discussed in Chapter 1 and addressed several measurement questions. Despite using increasingly robust designs, similar patterns have emerged for judgments of cause and blame across the five studies: the negative valence of the outcome is the primary influence on attributions of causality, whereas the valence of the agent's motive is the dominant influence for judgments of praise/blame. However, the most novel finding in these studies is that the degree of 'match' between the valence of the action and the outcome shapes people's causal attributions. For example, when the outcome 'matched' the agent's motive (i.e., negative motive and negative outcome; or positive motive and positive outcome), participants amplify the agent's causal contribution while discounting the proximal physical event. In contrast, when there was a 'mismatch' between motive and outcome (i.e., positive motive and negative outcome; or negative motive and positive outcome), participants discounted the agent's causal contribution in favour of proximal physical events. In contrast, for judgments of culpability the valence of the agent's motive explains people's attributions.

These findings are consistent with earlier research showing valence effects for judgments of causality and culpability (Alicke, 1992; Bohnet et al., 1988; Tetlock et al., 2007; Ybarra, 2002). The results also support several hypotheses (e.g., Alicke, 2000; Hart & Honoré, 1985) and the results for negative actions with negative outcomes replicate earlier research on causal chains (e.g., Hilton et al., 2010; Lagnado & Channon, 2008; McClure et al., 2007). However, none of these theories is able to account for the findings for causal chains that include positive actions or positive outcomes.

Applying theories of intentionality to causal chains

This thesis has examined the effects of positive and negative valence on judgments for intentional, rather than unintentional or accidental actions. Yet up until this point, I have not invoked theories of intentionality, such as that described by Malle and Knobe (1997), to explain the 'mismatch effect' found for judgments of cause. Malle and Knobe's (1997) theory of folk intentionality posits that people use distinct modes of explanation for intentional and unintentional actions. As with the covariation and social functionalist approaches discussed in Chapter 1, the model claims that *unintentional actions*, like physical events, are explained in a similar 'cause-effect' manner (Malle, 2011, p. 299). In contrast, *intentional actions* are explained by the agent's reasons for acting. In support of these claims, people attribute substantially more blame and causality when agent's acted intentionally than for unintentional actions or physical

events, even when the effects were the same (Hilton et al., 2010; Lagnado & Channon, 2008). Furthermore, unintentional actions were judged more negatively when they resulted in negative rather than positive outcomes, or harmful outcomes that could have been prevented or controlled (Alicke, 1992; Alicke et al., 2008; Ohtsubo, 2007; Weiner, 1995). For people to perceive an action as intentional, Malle and Knobe's model claims that five related components (beliefs, desires, intention, skill and awareness) must be present. Intention reflects the agent's mental state (or intent) at the time the action was performed, and is inferred when the agent desired and believed their action would bring about the effect. Intentionality, being a property of the action, is inferred when the agent has the intention, as well as skill and awareness when performing the action.

Summarising his work on intentionality, Malle (2011) suggested that the differences in attributions for intentional actions reflect the different types of reasons that people cite (e.g., the agent's beliefs and desires) which diverge depending on their perspective: the agent or the observer. For example, Malle, Knobe and Nelson (2007) found that when explaining intentional actions, agents provided more belief reasons whereas observers provided more desire reasons. Yet these differences were eliminated when the researchers manipulated the observer's perspective: when explaining actions from the agent's perspective, observers cited similar numbers of belief and desire reasons to agents. Malle suggested that this agent-observer asymmetry reflects differences in the relative ease that the differing types of reasons come to mind and pragmatic goals, such as making the agent look socially acceptable, or to deflect blame. Therefore, factors such as perspective and cognitive accessibility affect how people interpret and explain another's action.

Further evidence of asymmetries in attributions reflecting judgments of the agent's intent comes from research by Cushman and colleagues (Cushman, 2008; Young et al., 2007). In a study using functioning neuroimaging, Young et al. (2007) found that brain activity and moral judgments differed depending on the congruence between the agent's belief and the outcome. The greatest increase in activity was in regions associated with cognitive conflict when belief and outcome information were incongruent. Examples are when the action resulted in accidental harm (e.g., putting poison in a friend's coffee believing it to be sugar and the friend died) or the intended harm failed to eventuate (e.g., putting sugar in a friend's coffee believing it to be poison and the friend survived). However, this difference in activity did not map directly onto moral judgments. Instead, when harm was intended (e.g., putting poison in a coffee believing it to be

poison) regardless of the outcome (e.g., the friend died or survived) moral judgments were shaped by belief information, and by belief and outcome information when harm was not intended. These findings suggest that the congruence between the agent's intent and the outcome influences people's moral judgments.

Extending Young et al.'s (2007) design to include desire as well as belief, Cushman (2008) found that responses reflected the congruence between the intent, outcome and the type of attribution being made. Judgments regarding the wrongness of the agent's action were influenced by belief and desire reasons, but not the outcome. In contrast, judgments of blame depended mostly on beliefs and the outcome but not desire reasons. These results support Malle and colleagues' claim that people are sensitive to the different components of intentionality (Malle, 2011; Malle & Knobe, 1997; Malle et al., 2007). They also suggest that the relative importance of each component is malleable, depending upon the outcome, the type of question, and the observer's perspective (Cushman, 2008; Kashima et al., 1998; Malle et al., 2007).

It is plausible that the asymmetries in attributions reported in this thesis reflect participants judging the agent's intentions and intentionality in a manner that *explains* the decision to act than positive or negative valence effects *per se*. Indeed, participants were explicitly told that the agent intentionally performed the action. For example, the fire scenario read: '*A young man deliberately set fire to an area of forest. His goal was to cause a second fire (burn-off) to ensure (prevent) the wildfires reached (from reaching) the town*'. The implication here is that the action was performed with intent (the man *believes* and *wants* to start a fire), and he has the *ability* and *awareness* to perform the action which resulted in the intended outcome (a fire). In the matched conditions (e.g., positive action and outcome, or negative action and outcome), the agent's intent and the intended outcome are congruent with the scenario's outcome. Therefore, there is reason for participants to assume that the realised outcome was the intended outcome, so they infer intentionality, which in turn shapes their attributions. In contrast, in the mismatched conditions (e.g., positive action and negative outcome, or vice versa), the agent performs an action that fulfils the necessary components of intentionality, yet the sequence of events lead to an unintended outcome from the agent's perspective. Thus, participants could discount judgments of intentionality based on the realised outcome rather than the intended outcome, and support their assessment by weighing the intent factors (i.e., the agent's belief and desires) in a manner that is consistent with an outcome that was unintentionally brought about.

Intentionality and foresight from three different perspectives.

Drawing on Malle and Knobe's (1997) concept of intentionality, it is likely that awareness, being one element of the agent's epistemic state at the time of action, may capture a degree of foreseeability. Foresight, like intentionality, has long been seen as a key factor in lay attributions and legal reasoning (Harley, 2007; Nelson-le Gall, 1985). People tend not to blame or hold another person causally responsible for an outcome that they could not anticipate (Shaver, 1985). Similarly, the less foreseeable the outcome, the less likely it is that people see an action as intentional. However, a person may act intentionally, although the outcome was unforeseen. For example, a doctor might intentionally administer a drug knowing the side-effects were risky, but he did not intend to cause the patient's death.

Hindsight differs from foresight in temporal perspective as it is backward rather than forward-looking (Fischhoff, 1975)¹². Research has shown that the rater's point of view influences judgments of intentionality, cause, and blame (Menec & Weiner, 2000; Pezzo, 2011). Hindsight bias refers to people's tendency to overestimate the probability of the outcome occurring once the actual outcome is known. Similarly, people tend amplify information that is consistent with the outcome, but discount the same information when it is inconsistent with the outcome. One implication is that with hindsight, people may assume that the agent *should have known* that the outcome was the likely occurrence and weigh the agent's intent accordingly. Alternatively, they might infer, for example with negative outcomes, that the agent had hidden intentions and therefore intentionally (albeit covertly) facilitated the outcome.

Based on these reasons, the asymmetries in attributions reported in Studies 1-5 might reflect participants reconciling their perception of what they thought the agent expected the outcome to be when they acted (*subjective foresight*), with what was likely, regardless of what the agent expected (*objective foresight*), and what the agent should have known (*reasonable foresight*). For example, with the fire scenario, assume that 'the young man' acting with positive intent, had followed normal due diligence and believed that starting a burn-off was the correct course of action. Given that he did not expect a tornado to strike, the subjective foreseeability of a negative outcome is relatively low. However, a lawyer might argue that the man had not prepared adequately and should have identified important risk factors. Thus, the lawyer would argue that the outcome was reasonably foreseeable. Objective foresight, in contrast, reflects the participants'

¹² Hindsight conditions assess the probability of the outcome once the actual outcome is known whereas foresight conditions assess probabilities when the outcome information is withheld (e.g., Fischhoff, 1975).

assessment of whether the man's actions made a difference given that the town was cut off by wildfires and the residents were trapped in their homes.

Lagnado and Channon (2008) examined the relation between intentionality and two categories of foresight: objective foresight and subjective foresight in vignettes leading to negative outcomes. As predicted, participants rated actions as more causal and blameworthy when the outcome was foreseeable. However, the weightings that participants applied to the two categories of foresight differed depending on the attribution. Judgments of blame were influenced by both subjective and objective foresight regardless of the agent's intent; blame ratings were highest when the agent expected the adverse outcome (subjective foresight), and the outcome was indeed likely (objective foresight). Judgments of cause were also influenced by objective foresight regardless of the agent's intent; the more likely the adverse outcome was, the more the action was judged to be the cause. In contrast, subjective foresight reflected the agent's intent; the action was judged more causal when the expected outcome eventuated. However, judgments of cause did not differ when the agent's intent was not realised, or as described here, 'mismatched' the outcome.

Lagnado and Channon's (2008) results suggest that foresight reflects the observer's perspective rather than being a singular construct that influences judgments the same way regardless of perspective. Further, the findings imply that moral judgments of blame are influenced more by what people think is likely to occur, rather than what the agent thinks was going to occur. In contrast, when inferring the cause of an outcome, people appear to balance the causal factors in the chain from the agent's perspective with an objective analysis of what was likely to occur. The results show that observers are sensitive to the congruency between explicit statements of the agent's intent and what they expect will eventuate (subjective foresight). When there are subjective inconsistencies where belief information (i.e., what the agent believes is likely) differs from objective foresight (what in fact is likely), people discount the influence of both factors on judgments of cause. Yet, the findings also suggest that inconsistent belief and objective foresight information do not negate the influence of desire information (what the agent wants to occur) on judgments of blame, indeed, the central effect of desire on attributions in Lagnado and Channon's study is consistent with Cushman's (2008) findings with judgments of blame, permissibility and wrongness.

A New Attributional Theory: Sloman, Fernbach and Ewing's (2012) Causal Model

When considering the retrospective nature of attributions in these causal chains, it is possible that participants, with the benefit of hindsight, perceived two outcomes: the realised outcome as described in the scenario, and the agent's anticipated outcome. Thus, participants might assess the agent's intent in accordance the realised outcome rather than the intended outcome. The idea that multiple outcomes from a single event can have diverging effects on people's judgments has been the subject of debate since Knobe (2003) identified the 'the side-effect effect'. As discussed in Chapter 1, Knobe found that people's judgments of intentionality were sensitive not only to an action's intended outcome, but also to secondary outcomes or 'side-effects'¹³. Recently, Sloman et al., (2012) explained the side-effect effect by integrating Malle and Knobe's (1997) concept of intentionality with research on the morality, normativeness, and informativeness of the action (e.g., Machery, 2008; Nanay, 2010; Sripada, 2012; Uttich & Lombrozo, 2010), together with the notion of discounting (Kelley, 1973; McClure, 1998).

Sloman et al.'s (2012) causal model is based on four key premises: first, when multiple outcomes result from a single action, each consequence provides separate reasons for the action. Therefore, in cases of intended outcomes and foreseen side-effects, there are multiple associated belief and desire reasons. For example, in Knobe's (2003) chairman scenario, the agent's attitude towards the primary outcome (e.g. profit) and the side-effect (e.g., helping or harming the environment) results in two belief and desire nodes: one for the primary outcome and the other for the side-effect. Second, as in the culpable control model (Alicke 2000), Sloman et al. suggest that people's judgments reflect their evaluative reactions to the agent and to the events. To demonstrate these claims, Sloman et al. extended Malle and Knobe's (1997) model of intentionality by adding character and consequence components. This adaptation provided a conceptual framework for Sloman et al.'s core proposal that people use two inferential pathways to assess the intentionality of another's action: a bottom-up inferential mechanism based on factual information about the consequences, the action, and alternative causes, and a top-down inferential mechanism based on information about the agent's character and likely disposition towards the outcome¹⁴.

¹³ See Chapter 1, p. 14 for further discussion and illustration of the side-effect effect using Knobe's (2003) chairman scenario.

¹⁴ Sloman et al. (2012, p. 160) use the term 'inferential mechanism' to describe the process people use to infer another's attitude towards the side-effect.

Third, Sloman et al. (2012) suggest that judgments of intentionality depend not only on what the observer estimates the agent's intent to be, but also on what the outcome might have been if the agent's intent had been different. The fourth premise is that when several causes independently contribute to the outcome, one cause is discounted in favour of another. However, the following conditions are necessary for discounting: Firstly, the side-effect provides a normative reason *to act*, but not when it opposes action. Secondly, when there are several independent reasons for the action, people discount those reasons that add little informative value or do not provide the necessary reasons for action.

Applying these claims to Knobe's (2003) chairman scenario, for example; there are two independent reasons for the action: the primary goal (profit), being the same in both the help and harm conditions, and the second reason associated solely with the side-effect. In the negative condition, the side-effect of harming the environment provides a normative reason *not to act*. So, if the chairman had cared about harming the environment, the outcome might have been different. The inference then is that the chairman's failure to avoid harm, tells us about his intent as he knowingly allowed the side-effect to occur, regardless of his desire for the primary reason (e.g., to make a profit). Thus, his action becomes the necessary condition so the assumption of independence is violated, and people judge him as having intentionally harmed the environment. In contrast, in the positive side-effect condition there are two sufficient reasons for acting; making profit and helping the environment. Therefore, his attitude towards the side-effect is a sufficient, but not necessary reason for acting. Thus, the independence criterion is not violated and people's judgments of intentionality for the side-effect are discounted in favour of the more informative primary reason to act, to make profit. Sloman et al. supported the model with experiments that compared participants' judgments for the two outcomes: the intended outcome and the side-effect. The predicted discounting patterns emerged for judgments about intentionality depending on the match between the normative reasons for the action and the valence of the side-effect. This finding suggests that extending Sloman et al.'s (2012) model to these causal chains with intended foreseen and unforeseen outcomes, might explain the mismatch effects reported in these studies.

Taken together, these theories and findings suggest that lay attributions of cause and blame depend not only on the core components of human agency, intentionality and foresight, but also on people's perspective, and the extent that the events depart from normative expectations. This raises two interesting questions. For example, how does the valence of the motive or outcome

interact with the intentional properties of an agent's action to influence people's judgments of multiple events in causal chains? How does the observer's perspective interact with judgments of intentionality to guide judgments of cause and blame?

Study 6: Testing the Role of Intentionality and Foresight

Overview of the Research Hypotheses

Study 6 was designed to address these questions by investigating whether people's perception of intentionality and foresight predicted the effects of valence on judgments of cause. Several predictions were based on previous research that considered similar questions examining attributions about a single target event in the chain (e.g., Cushman, 2008; Kashima et al., 1998; Malle & Knobe, 1997; Malle et al., 2007) and Sloman et al.'s (2012) causal model. The first prediction was that judgments of the agent's intent (belief and desire), intentionality, and the three types of foresight (subjective, objective, and reasonable foresight) would differ depending on the match between the valence of the motive and the outcome. Specifically, I predicted a 2-way interaction between the valence of the outcome and the type of distal event where ratings of intent (belief and desire) and intentionality would be higher in matched than mismatched conditions. These predictions test the notion that when the motive and outcome are mismatched, people evaluate each component of intentionality in a manner that is consistent with the realised outcome than the agent's intended outcome.

The second prediction was that judgments of foresight (subjective, objective, and reasonable foresight) would differ depending on the match between the valence of the motive and the outcome. I predicted that ratings of subjective foresight would be higher than objective foresight in the matched than mismatched conditions. Specifically, I predicted that participants would amplify ratings of foreseeability when the anticipated outcome eventuated, but attenuate ratings when an unanticipated eventuated when considering the events from the agent's perspective (subjective foresight) than an objective perspective. This prediction tests the notion that differences in the outcome's foreseeability reflect the explainer's perspective (Malle, 2011). In contrast, reasonable foreseeability concerns what the agent *should have* expected and reflects hindsight bias. Consequently, as participants consider the alternative actions the agent could have performed, the congruence between the agent's motive and outcome has a diverging effect on ratings of reasonable foresight. Therefore, I predicted that ratings of reasonable foresight would be higher than objective foresight in the matched conditions. In contrast, ratings of reasonable and objective foresight would be higher than subjective foresight in the mismatched conditions.

Negative outcomes however, would amplify ratings of reasonable foresight but not objective or subjective foresight. This prediction tests Malle's (2011) claim that judgments of reasonable foresight are not constrained by pragmatic goals (e.g., social desirability or to deflect blame). In statistical terms, this predicts a 3-way interaction between the valence of the outcome, the type of distal event, and foresight (subjective, objective, and reasonable) for ratings of the distal event.

Several predictions were derived from Sloman et al.'s (2012) model to see whether their explanation of the side-effect effect could account for the valence effects in these studies. The bottom-up inferential process supports two predictions: first, with negative outcomes, people will discount the causal effects of positive actions in favour of alternative explanations (i.e., the proximal physical event) but not negative actions because the anticipated outcome provides a normative reason for the agent to act when their goal was to prevent harm but not to ensure harm eventuates. In contrast, in chains with positive outcomes, the agent's causal contribution will not be discounted in favour of an alternative explanation (the proximal event) because a change in the agent's intent will not result in a better outcome. Second, judgments of intentionality and foresight will be mitigated when another known independent event (e.g., proximal physical event) provides a more informative explanation of the realised outcome. In addition, the top-down inferential predicts that causal ratings will be amplified when actions violate perceived norms or produces harm.

Based on this approach and incorporating observer and perspective biases, I predicted that for chains with negative outcomes, participants would see negative actions as intentional, the outcome as foreseen, and would discount subsequent physical events. The reverse pattern would emerge for positive actions: participants would see positive actions as unintentional, the outcome as unforeseen, and discount the actions in favour of subsequent physical events. In contrast, for chains with positive outcomes, participants would judge positive and negative actions and physical events as equal causes of the outcome. In addition, participants would see positive actions as intentional and the outcome as foreseen, but negative actions as unintentional and the outcome as unforeseen. Further, participants will see the proximal event as the better explanation and amplify causal ratings when the agent's intent violates normative action.

Thus in statistical terms, I predicted two 3-way interactions: I predicted a 3-way interaction between outcome, motive, and temporal position for judgments of cause, and a 3-way interaction between outcome, motive, and type of judgment (intent, intentionality, and the three types of foresight) for ratings of the distal action.

Overview of the Experimental Paradigm

One fundamental change was made to the design used in Study 5. The distal physical event in Studies 1-5 was replaced by an unstated motive, for two reasons. First, an unstated motive provides a baseline to address the question of whether the valence of the motive rather than intentionality *per se* affects attributions of cause and praise/blame. Second, it allows a test of Sloman et al.'s (2012) claim that people consider not only how much the agent wanted (desired) the outcome, but also what might have been if the agent did not hold that desire.

To investigate whether Sloman et al.'s (2012) model accounts for the effects of valence on judgments of cause and blame, the components of intent (assessed by two items: belief and desire) and intentionality were added as variables. The decision to assess belief and desire independently rather than as a single measure of intent permits tests of specific predictions on the distinctions people draw between judgments of cause and blame (Cushman, 2008; Malle et al., 2007). The components of skill and awareness have important roles in intentional actions, but were not assessed because in these scenarios they have the same effect across all conditions. For example, given the successful production of the intended primary outcome (*e.g., a fire is started*), it is assumed that the man had the *skill* and *awareness* to perform the action.

To investigate whether the perspective people take when evaluating another's actions can account for the effects of valence on judgments of cause and blame, participants rated the foreseeability of the outcome from three perspectives; the agent's point of view (subjective foresight), their own view of what was likely to eventuate given the presence or absence of the preceding events (objective foresight) and what the agent should have expected (reasonable foresight).

Heider's (1958) concept of equifinality suggests that judging an action as intentional results in people seeing the chain as a more integrated sequence of events. However, judgments of the unitary nature of the chain in Studies 1 and 2 did not support this hypothesis. This finding may be due to the fact that judgments of intentionality mediated the relations between the type of action and causality. Because intentionality was being measured in Study 6, I reintroduced the unitisation measure. The inconsistent findings with this measure in research by McClure et al., (2007), Hilton and McClure (2013) and in Studies 1-2 may reflect an ambiguous measure and anchors. To clarify the measure, the measurement and anchor descriptions were reworded.

To summarise, the distal physical event condition in Studies 1-5 was replaced with a third distal action condition, where information regarding the agent's motive was not stated. To

examine the influence of participants' perception of intentionality and foreseeability on their judgments of cause and praise/blame, I added measures of intent (belief and desire), intentionality, and foresight. The measure of responsibility used in Studies 2-4 was reintroduced to examine whether intentionality and foresight have distinct effects on judgments of cause, praise/blame and responsibility. The measure of the unitisation of chains used in Studies 1-2 was revised and examined whether this measure predicts judgments of cause.

Method

Design.

As in Study 5, Study 6 used a mixed design that included three between-subject variables: scenario (fire, train, factory, avalanche, car, casino, or soldier), outcome (positive or negative) and distal event (positive action, negative action, or unstated motive action)¹⁵ and a fourth within-subject variable that was the temporal position of the causal event (distal or proximal). The 2 (outcome) x 3 (distal event) design resulted in six versions of each scenario.

The action was the same in all motive conditions. For example, in the fire scenario the action read: *'A young man deliberately set fire to an area of forest'*. The action was then followed by either a positive motive e.g., *'His goal was to cause a burn-off to prevent the wildfires from reaching the town'*, a negative motive e.g., *'His goal was to cause a second fire to ensure the wildfires reached the town'*, or the new condition with no motive information.

Scenarios.

One change was made to all scenarios. The outcome statement was revised to read *'As a result'* because the statement in the previous studies *'These events resulted in'* implies that both events have a causal connection to the outcome, potentially biasing perceptions of the unitary nature of the causal chain. Minor revisions were made to several scenarios to ensure structural equivalence across scenarios (Appendix A, pp. 148-159).

Measures and procedure.

Participants ($N = 415$) read a single scenario then gave judgments of the distal and proximal events' causality, praise/blame, responsibility, intention, foreseeability (subjective, objective, and reasonable foresight), and unitisation of chain. The measures were counterbalanced in a Latin square design. Mandel's (2003) probability task was presented as either the first or last item on the questionnaire; the order was reversed for half of the conditions.

¹⁵ Distal event is hereafter referred to by the 3 motive conditions: positive, negative, or unstated.

The measures of cause, praise/blame and probability were same as Study 5, and the measure of responsibility was the same as Studies 2, 3 and 4. The measures of intent and intentionality were those used by Guglielmo and Malle (2010b). The measure of intent assessed belief and desire using a 2-item scale. The belief item read: *To what extent do you think the man thought his actions would result in the outcome in this scenario?* The desire item read: *To what extent do you think the man wanted the outcome in this scenario?* These two items were rated on an 11-point scale ranging from 0 = Not at all to 10 = Absolutely. The measure of intentionality read: *To what extent do you think the man intentionally produced the outcome in this scenario?* Ratings were on a scale, where 0 = Not at all intentional to 10 = Completely intentional.

The measures of objective and reasonable foresight used Mandel's (2003) probability task. Objective foresight was assessed as the probabilistic contribution of the distal action, to reflect the degree to which the action raised the probability of the outcome occurring [$P(O|n+d)$ - $P(O|n)$]. Reasonable foresight was assessed as the likelihood of the outcome after the distal action occurred, given the information available to the agent at the time [$P(O|n+d)$]. The measure of subjective foresight used Menec and Weiner's (2000) two-item foreseeability measure: *To what extent do you think the man could foresee [predict] the outcome in this scenario?* Each item was rated on an 11-point scale ranging from 0 = Not at all foreseeable [predictable] to 10 = 'Completely foreseeable [predictable]'. Replicating Menec and Weiner's results, the two items were highly correlated (coefficient alpha .72) and were combined to give a mean subjective foresight score.

The unitisation measure from Studies 1 - 2 was revised to read: 'Do you think the two events described above seem like two separate incidents or a single integrated sequence that resulted in the outcome?' Ratings were on a bipolar scale with 0 as the neutral midpoint, and -5 = Two separate incidents and 5 = A single integrated sequence.

Results

Targeted analyses of the baseline condition with unstated motive showed that scenario did not interact with any other factors, indicating that any effects of valence cannot be explained by differences between scenario conditions, so subsequent analyses collapsed this factor.

To examine whether the effects in Studies 1-4 were replicated in Study 5, a mixed design ANOVA was conducted with two between-subjects variables: 2 (outcome: positive, negative) x 3 (motive: positive, negative, unstated) and two within-subjects variables: 2 (temporal position: distal, proximal) x 3 (judgment: cause, responsibility, praise/blame). See Figure 5.1 and

Appendix J1 for descriptive statistics. The analyses revealed main effects for outcome, $F(1, 409) = 27.42, p < .001, \eta^2 = .06$, and motive, $F(2, 409) = 11.43, p < .001, \eta^2 = .05$, but not for type of judgment or temporal position. Importantly, the main effects were qualified by a 4-way interaction between outcome, motive, temporal position, and judgment, $F(4, 818) = 5.70, p < .001, \eta^2 = .03$.

Figure 5.1. Means of cause and praise/blame for actions leading to positive and negative outcomes with three motive conditions (Study 6).

Influence of motive, and outcome on attributions.

To clarify the 4-interaction between outcome, motive, temporal position and type of judgment, separate MANOVAs with between-subjects variables: 2 (outcome) and 3 (motive) were conducted for judgments of cause and praise/blame¹⁶.

¹⁶ Mixed design ANOVA was conducted with two between-subjects variables: 2 (outcome: positive, negative) x 2 (motive: positive, negative or unstated), and two within-subjects variables: 2 (temporal position: distal or proximal) x 2 (judgment: cause and responsibility). The analysis revealed a 3-way interaction between outcome, motive and temporal position, $F(2, 409) = 19.57, p < .001, \eta^2 = .09$, but not

Cause judgments.

With judgments of cause, replicating Study 5, there was a multivariate 2-way interaction between outcome and motive, $F(4, 818) = 7.08, p < .001, \eta^2 = .03$, but no main effects for outcome or motive. Follow-up univariate analyses showed interactions for ratings of distal actions, $F(2, 409) = 9.63, p < .001, \eta^2 = .05$, and proximal events, $F(2, 409) = 8.56, p < .001, \eta^2 = .04$, so separate analyses were conducted for positive and negative outcomes.

For negative outcomes, as in earlier studies, there was a main effect for motive for distal actions, $F(2, 204) = 4.29, p < .05$, and proximal events, $F(2, 204) = 4.29, p < .05$. Negative actions were rated more causal than positive and unstated actions, whereas the reverse pattern was shown for proximal physical events. Post hoc Bonferroni ($p < .05$) tests showed that physical events that followed positive and unstated actions were rated higher than those following negative actions. For positive outcomes, as in earlier studies, the valence of the agent's motive influenced causal ratings for proximal events, $F(2, 204) = 4.29, p < .05$, but not distal actions. Proximal physical events were rated higher when following negative actions than positive actions. Physical events that followed unstated actions did not differ from either motive condition.

Independent *t*-tests examined whether the predictions based on Sloman et al.'s (2012) bottom-up inferential process could account for the valence effects on judgments of cause. When comparing the three distal actions across positive and negative outcomes, negative outcomes augmented ratings of negative actions, $t(137) = 2.64, p < .01$, attenuated ratings of positive actions, $t(138) = 3.86, p < .001$, but had no effect on ratings for unstated motive actions. In contrast, negative outcomes attenuated ratings of proximal events when following negative actions, $t(137) = 3.78, p < .001$, but had no effect on proximal events following positive actions or unstated actions. When comparing the influence of motive on ratings for both events, with positive outcomes, proximal physical events were rated *more causal* than unstated actions, $t(67) = 2.59, p < .01$, and negative actions, $t(69) = 3.35, p < .001$, but *as causal* as positive actions. In contrast, for negative outcomes, proximal physical events were rated *more causal* than positive actions, $t(69) = 4.14, p < .001$, and unstated actions, $t(67) = 2.59, p < .02$, but *less causal* than negative actions, $t(69) = 3.35, p < .001$.

a 4-way interaction with judgments, $F(2, 409) = 0.81, ns$, so results hereafter are reported for judgments of cause only.

Overall, these results support Sloman et al. (2012) proposal that there are two inferential processes shaping judgments of cause as the predicted patterns of discounting and amplification emerged in five of six conditions. As predicted for positive outcomes, regardless of the presence or valence of the agent's motive, participants did not discount the agent's causal effect when there was no normative reason to change the positive outcome. Further, as predicted participants did amplify causal ratings for proximal events when preceded by negative actions. With negative outcomes, the predicted effects were found when the agent's motive was stated but not unstated. Specifically, as predicted, participants discounted the agent's causal effect in favour of proximal physical events, which in turn were further amplified when there was good reason to change the outcome but the positive action adhered to normative standards. In contrast, participants discounted the proximal physical events causal effect in favour of negative actions, which in turn were also amplified, as there was a good to change the negative outcome and the action violated social norms. However, in contrast to predictions, when the motive was unstated the patterns replicated those for positive not negative actions.

Praise/Blame judgments.

With praise/blame judgments, replicating Study 5, there was a multivariate main effect for outcome, $F(4, 824) = 26.73, p < .001, \eta^2 = .03$, and motive, $F(4, 824) = 26.73, p < .001, \eta^2 = .03$, but no interaction, see Figure 5.1. Follow-up univariate analyses showed a main effect for outcome for distal actions, $F(1, 409) = 47.37, p < .001, \eta^2 = .10$, and proximal events, $F(1, 409) = 140.76, p < .001, \eta^2 = .26$, where both events were rated higher with negative outcomes. The analyses also showed a main effect for motive for ratings of distal actions, $F(2, 409) = 61.12, p < .001, \eta^2 = .23$, and proximal events, $F(2, 409) = 4.52, p < .05, \eta^2 = .04$. Negative actions were rated higher than unstated actions, which in turn were rated higher than positive actions. The valence of the agent's motive had the reverse effect on judgments of proximal events. Thus, physical events were rated higher when following positive rather than negative actions but did not differ for unstated actions.

In sum, judgments of cause and praise/blame with positive and negative actions replicate those found in Studies 1-5. Interestingly, when the motive was unstated, the patterns differed depending on the attribution. For judgments of cause, unstated actions resemble positive actions, whereas, for judgments of praise/blame the patterns reflect the valence of the outcome; participants judged that the agent's action blameworthy with negative outcomes, but neither praiseworthy nor blameworthy with positive outcomes.

Influence of motive and outcome valence on judgments of intentionality and intent.

To test whether the valence of the agent's motive or outcome influenced judgments of intentionality and intent (belief and desire), a MANOVA was conducted with two between-subjects variables: 2 (outcome) and 3 (motive) for judgments of intentionality, belief, and desire. See Figure 5.2 and Appendix J2 for descriptive statistics. Supporting predictions drawn from research (e.g., Malle et al., 2007) and Sloman et al.'s (2012) model, there was a multivariate 2-way interaction between outcome and motive, $F(4, 818) = 15.54, p < .001, \eta^2 = .07$. Follow-up univariate analyses showed interactions for judgments of intentionality, $F(2, 409) = 43.75, p < .001, \eta^2 = .18$, belief, $F(2, 409) = 55.54, p < .001, \eta^2 = .21$, and desire, $F(2, 409) = 142.94, p < .001, \eta^2 = .41$.

Figure 5.2. Mean ratings for the intent (belief and desire) and intentionality for actions leading to positive and negative outcomes with three motive conditions (Study 6).

To test whether judgments of intent (belief and desire) and intentionality follow the mismatch effect, separate MANOVAs were performed for each motive condition. As predicted, the mismatch effect emerged: the intentionality of positive actions and unstated actions was rated higher in chains with positive than negative outcomes (positive actions $F(3, 136) = 62.04, p < .001$, unstated actions, $F(3, 132) = 25.35, p < .001$), whereas negative actions were rated higher in chains with negative than positive outcomes, $F(3, 135) = 26.89, p < .001$.

These results show that the valence of the agent's motive and the outcome influences judgments about the agent's intent and whether the agent intentionality brought about the outcome. Interesting too, is the finding that participants did not evaluate the agent's intent or

intentionality as explicitly stated in the scenarios. Instead, they evaluate intent in a manner that reflects the degree of congruence between the agent's motive and the realised consequences.

Influence of motive and outcome valence on judgments of foreseeability.

To examine whether the valence of the outcome and motive influenced judgments of foresight, a mixed ANOVA was conducted with the between-subjects variables: 2 (outcome) and 3 (motive), and the within-subjects variable 3 (foresight type: objective, subjective or reasonable). See Figure 5.3 and Appendix J3 and for descriptive statistics. This analysis revealed main effects for outcome, $F(1, 409) = 7.81, p < .05, \eta^2 = .02$, and foresight, $F(2, 818) = 258.78, p < .001, \eta^2 = .39$. Importantly, there was a 3-way interaction between outcome, motive, and type of foresight, $F(4, 818) = 3.73, p < .01, \eta^2 = .02$.

Figure 5.3. Mean ratings for subjective, objective and reasonable foresight by condition. Error bars show standard errors (Study 6).

To clarify the influence of the outcome on judgments of foreseeability, separate mixed ANOVAs were performed for each of the three motive conditions. For chains with negative actions there were main effects for outcome, $F(1, 137) = 70.83, p < .001, \eta^2 = .34$, and foreseeability, $F(2, 274) = 81.64, p < .001, \eta^2 = .37$. The main effects were qualified by a 2-way interaction, $F(2, 136) = 8.95, p < .001, \eta^2 = .12$. For each type of foresight, negative outcomes were more foreseeable than positive outcomes. With negative outcomes, participants judged that

the agent foresaw the harm, that the harm was indeed likely to eventuate, and that the agent should have expected the negative outcome. In contrast, with positive outcomes, participants considered it reasonable to expect a good outcome, but they did not think it was likely to occur or that the agent had foreseen it. In fact, the results suggest that participants judged the agent's negative motive negated the likelihood that good would eventuate.

For chains with positive actions, there was a main effect for outcome, where positive outcomes were rated more foreseeable than negative outcomes, $F(1, 138) = 29.66, p < .001, \eta^2 = .18$. There was a main effect for foreseeability, where ratings were higher for reasonable than subjective or objective foresight, which did not differ, $F(2, 276) = 83.07, p < .001, \eta^2 = .38$. There was no interaction between outcome and foresight. For chains with unstated motives, there was a main effect for foreseeability, where ratings were higher for reasonable than subjective foresight, which in turn was higher than objective foresight, $F(2, 133) = 158.02, p < .001, \eta^2 = .70$. There was no main effect for outcome, or interaction between outcome and foresight.

In sum, contrasts showed that participants do distinguish between the three types of foresight: judgments of reasonable foresight were higher than judgments of subjective foresight, which in turn were higher than judgments of objective foresight. Importantly, these findings show the match between the agents motive influences people's judgments of foreseeability. With positive actions, positive outcomes were more foreseeable than negative outcomes, whereas with negative actions, negative outcomes were more foreseeable. When the motive was unstated, the valence of the outcome did not influence judgments of foreseeability. This finding demonstrate that motives shape people's assessment of an outcome's foreseeability; negative motives amplify foresight ratings beyond where they reside if the motive was unknown, whereas with positive motives, people discount the outcome's foreseeability. Thus, as predicted, these findings show that judgments of the outcome's foreseeability systematically vary depending on the perspective people take and the congruence between the valence of the motive and outcome. The outcome is foreseeable when the valenced conditions are matched, but not when they are mismatched.

Tests of Two Theories of Intentionality

Heider's (1958) concept of equifinality.

Two sets of covariance analyses were conducted with between-subjects variables: 2 (outcome) and 3 (motive) using the difference score between the proximal and distal events. See Appendix J2 for descriptive statistics. The first set of analyses tested if intentionality influenced judgments of the unitisation, and the second set tested if unitisation predicted causal judgments.

First, the analysis removed intentionality ratings from the interaction between outcome and motive for judgments of unitisation, $F(2, 409) = 8.92, p < .001, \eta^2 = .04$. Counter to predictions, judgments of intentionality did not influence judgments of the unitary nature of the causal chain. The second set of analyses removed unitisation ratings from the interaction between outcome and motive for judgments of cause. The analysis showed that judgments of unitisation had a small effect ($b = .10$) on judgments of cause, $F(1, 137) = 4.94, p < .03$. However, the interaction remained significant after covarying out the influence of unitisation on causal judgments, $F(2, 408) = 13.80, p < .001, \eta^2 = .06$. These findings suggest that how people perceive the links between two independent events does influence their causal attributions, but it is the ‘match’ between the agent’s motive and the outcome rather than intentionality *per se* that shapes their perception of the unitary nature of the chain.

Testing Malle and Knobe’s (1997) folk concept of intentionality.

Two sets of covariance analyses were conducted with between-subjects variables: 2 (outcome) and 3 (motive) tested how well Malle and Knobe’s (1997) components of intention (belief and desire) and foresight explained judgments of intentionality. The first set of analyses statistically removed belief and desire ratings from the main effect for motive and the interaction between outcome and motive on intentionality ratings. The results showed that intent (desire, $b = .41$, belief, $b = .27$) explained a significant portion of people’s judgments of intentionality (desire, $F(2, 407) = 36.23, p < .001$, belief, $F(2, 407) = 26.45, p < .001$). After covarying out ratings of belief and desire on judgments of intentionality, the main effect for motive remained significant, $F(2, 407) = 6.18, p < .01$, but the interaction between outcome and distal event did not.

The second set of analyses statistically removed the three types of foresight from the main effect for motive and the interaction between outcome and motive on intentionality ratings. The results showed that subjective foresight ($b = .58$) alone influenced judgments of intentionality, $F(2, 409) = 151.64, p < .001$. After covarying out ratings of foresight on judgments of intentionality, the interaction between outcome and motive remained significant, $F(2, 409) = 8.28, p < .001$, but the main effect for motive became insignificant.

Extending the folk concept of intentionality to include three categories of foresight.

Multiple regression analysis tested how well the combined properties of intent, intentionality and foresight explain judgments of cause and praise/blame. No evidence of multicollinearity was found.

For judgments of cause, the model explained 17% of the variability in ratings, $R = .41$, $F(9, 405) = 8.96$, $p < .001$, see Figure 5.4. Judgments of whether the outcome was likely (objective foresight, $b = .36$) had the strongest influence on judgments of cause. Inferences regarding whether the agent foresaw and intended to cause the outcome, provide a second tier of equally important predictors (intentionality, $b = .17$; subjective foresight, $b = .16$). Whether participants believed the outcome should have been foreseen (reasonable foresight, $b = .06$) and the agent's intent did not directly influence judgments of cause. Finally, the valence of the motive, and in particular negativity, had a small ($b = .14$) effect on judgments of cause, but the valence of the outcome did not.

For judgments of praise/blame, the model explained 34% of the variability in ratings, $R = .59$, $F(9, 405) = 23.33$, $p < .001$, see Figure 5.5. In contrast to causality, the presence rather than absence of motive ($b = .46$) and the valence of the outcome ($b = .23$) had the largest effect on judgments. Furthermore, a positive motive amplified judgments of praise ($b = -.24$), whereas a negative motive amplified judgments of blame ($b = .29$). Judgments of whether the agent believed their action would bring about the outcome (belief, $b = -.16$), and belief that the agent should have foreseen the outcome (reasonable foresight, $b = .17$), provide a second tier of equally important predictors. Judgments of whether the agent foresaw, wanted, or intentionally brought about the outcome, and whether participants thought the outcome was likely did not influence judgments of praise/blame (subjective foresight; $b = .08^{\text{ns}}$; objective foresight, $b = .09^{\text{ns}}$; desire; $b = .08^{\text{ns}}$; intentionality, $b = .00$).

In sum, these results show that each type of foresight plays a different role in judgments of intentionality, cause, and praise/blame. Subjective foresight influenced judgments of intentionality, whereas objective foresight had strongest influence on judgments of cause, and reasonable foresight had the stronger effect on judgments of praise/blame. These results replicate Lagnado and Channon's (2008) finding that objective probabilities had the stronger effect on judgments of cause but not their finding that subjective foresight had the stronger effect on judgments of praise/blame. These findings suggest that attributions are influenced by the observer's perspective. When attributing cause, participants take an objective perspective and

Figure 5.4. Causal Model: Regression Analyses (using dummy variables for motive): $R = .41^{**}$

Figure 5.5: Praise/Blame Model: Regression Analyses (using dummy variables for motive): $R = .59^{**}$

evaluate what in fact was likely to occur, in addition to assessing intentionality. Overall however, their judgments of causality, and components thereof, were primarily shaped by the match between the agent's motive and the outcome. In contrast, when attributing praise/blame, their judgments are influenced by what they think the agent should have foreseen, and their perception of what the agent believed would eventuate rather than intent and intentionality *per se*. Again, the valence of the agent's motive has the strongest effect on judgments of praise/blame.

Discussion

Tests of the patterns for outcome and motive in judgments of cause and praise/blame in Study 6 showed effect sizes similar to the previous five studies. Judgments of cause and responsibility followed the same pattern wherein ratings were influenced by the valence of the agent's motive and outcome. In contrast, judgments of praise/blame primarily reflected the agent's motive. Study 6 also investigated whether theories of intentionality (e.g., Malle & Knobe, 1997; Sloman et al., 2012) or the participants' perspective (e.g., Malle, 2011) could account for these effects and addresses several other questions. Each will be discussed in turn.

The role of intentionality in attributions.

According to Malle and Knobe's (1997) theory of intentionality, people's inferences about the agent's intentions are based on belief and desire reasons, which together with an evaluation of the agent's skills and awareness at the time of action, influence their assessment of intentionality. The results from this study support this view: participants judged an outcome as intentional when they thought the agent believed and wanted that effect, but saw it as unintentional when they judged the agent neither believed nor wanted that effect. Furthermore, consistent with Malle et al.'s (2007) finding that observers use more desire reasons and fewer belief reasons than actors do when explaining actions, participants' judgments of intentionality were influenced more by desire information than by belief information. However, the results in Study 6 extend Malle and colleagues' research to suggest that the framing of the agent's motive and the outcome, in addition to the rater's perspective, has a major effect on how people perceive another's intent.

A key novel finding in this study is that participants' judgments of intent (belief and desire) and intentionality reflect a mismatch effect, where these ratings depend on the congruence between the valence of the agent's motive and the outcome. Indeed, as found for judgments of cause and responsibility in Studies 1-5, the extent of match between the valence of the agent's motive and the realised outcome shapes participants' judgments of intent and intentionality.

Sloman et al. (2012) proposed that people's inferences of intentionality follow a bottom-up mechanism from effect to cause. Applying this claim to the present study, I predicted that based on the description of the action and the successful production of the intended outcome (*e.g., a fire is started*), participants would infer that the agent had acted with intent. Further, I predicted that participants would infer intent regardless of whether motive information was present or withheld in the baseline condition. Yet in this study, it is only when the valence of the motive for the intentional action matches the valence of the realised outcome (i.e., positive actions preceding positive outcomes, or negative actions preceding negative outcomes) that participants judge that the agent wanted and believed their action would bring about the outcome, and that the outcome was produced intentionally. In contrast, in sequences where the valence of the motive and outcome are mismatched, participants discount the agent's primary intent and the intentionality of the action by judging that the agent neither wanted nor believed that their action would produce the outcome, and that the outcome was unintended.

Further, the findings in this study show that these asymmetries in participants' judgments cannot be accounted for by the pervasive effect of negativity *per se*. Comparing judgments of intent and intentionality across the outcome conditions when the motive was not stated, participants did not assume intent or intentionality based on the action description. Instead, participants inferred that the agent had acted with positive intent as their judgments replicated that found for positive actions regardless of the outcome. For example, when the outcome was positive, the pattern of results for the unstated motive replicates that for the positive motive; as predicted, participants judged the agent wanted and believed their action would bring about the outcome, and that the outcome was intentional. Counter to predictions, they did not infer negative intent when the same action was intentionally performed but harm eventuated. Instead, their responses were more closely aligned with positive actions that went wrong; the agent was judged as neither wanting, believing, nor intending harm. This finding suggests that in the absence of evidence to the contrary, people assume that others mean well and that their action follows social norms rather than being shaped by what eventuated (Ybarra, 2002).

The results in Study 6 therefore suggests that people's inferences regarding another's action are malleable in that they weigh the components of belief and desire in a manner that supports their own evaluation of whether the agent acted intentionally rather than reflecting the agent's stated intention when the action was performed. Indeed, these findings support Malle (2011) and Kashima et al.'s (1998) explanations of why attribution theories that lay people employ differ

from those proposed by psychological or legal theorists. Specifically, the authors suggested that in order to understand another's action, people compare the action with an ideal prototype of the beliefs, desires, and intentions that correspond to normal action. The extent that the agent's action matches or deviates from the prototype shapes their assessment of the agent's intent, and in turn, whether the outcome was intentionally produced. Hence, when the story describes the agent's beliefs, desires, and intentions that match the ideal case, people judge the action as intentional and weigh the components of intentionality accordingly. But when the outcome was not anticipated or the action violates the ideal or prototypical action, the components of belief and desire are discounted and the action is judged unintentional.

The relations between judgments of intentionality, cause, and praise/blame.

The second novel finding in this study is that people see intent and intentionality differently when ascribing cause and praise/blame. Indeed, the results show that both components are necessary for judgments of cause, but not for judgments of praise/blame. Specifically, supporting Malle and Knobe's (1997) hierarchical model of intentionality, the results show that judgments of intent (belief and desire) predicted judgments of intentionality, which predicted participants' judgments of cause. Notably, judgments of intent did not predict judgments of cause. Yet, contrary to predictions, belief was the only component of intentionality that was directly related to judgments of praise/blame. Indeed, these results are more consistent with Cushman's (2008) findings where judgments of the agent's belief rather than desire or intentionality had the stronger relations with judgments of praise/blame. Thus, the present results suggest that people integrate the components of intent and intentionality for judgments of cause but treat them independently when judging praise/blame.

The influence of foresight: predicting intentionality, cause, and praise/blame.

The results in Study 6 are consistent with the hypothesis that the foreseeability of the outcome depends on the rater's perspective. Participants considered that the outcome in these scenarios, was reasonably foreseeable and that the agent should have foreseen it, even though the outcome was not foreseen when the action occurred (subjective foresight), or thought it likely to occur (objective foresight). The patterns for each perspective, however, show the same mismatch effect as reported for judgments of cause and intentionality. The outcome was foreseen from all perspectives when the conditions were matched but not when they were mismatched. Positive outcomes were foreseeable following positive actions as were negative outcomes following negative actions. In contrast, when the actions and outcomes were mismatched, the outcome was

less foreseeable, but ratings differed depending on perspective. Although participants considered the mismatched outcome was unintentional and indeed unforeseeable from their own perspective and the agent's perspective, the agent nevertheless 'should have' foreseen the outcome. Interestingly, the valence of the outcome did not influence judgments of foresight for unstated motives; here the outcome was both reasonably foreseeable and foreseen when the agent acted, although participants judge that the agent's action did not contribute to the foreseeability of the outcome. This suggests that the presence and valence of the agent's motive influences people's perception of an outcome's foreseeability.

Complementing Malle et al.'s (2007) work on actor-observer asymmetries in behavioural explanations; these results show that the participant's perspective has a similar asymmetrical effect on judgments of intent, intentionality, cause and culpability. Participants' judgments of the agent's intent and the intentionality of the outcome reflected their perceptions of how foreseeable the outcome was from the agent's perspective. In contrast, judgments of cause and praise/blame reflected the foreseeability of the outcome from the observer's perspective. This finding makes sense if considered in terms of the different functional roles that intentionality, cause, and praise/blame play in lay reasoning. Judgments of intention and intentionality rely more on the agent's epistemic state at the time, whereas judgments of cause and praise/blame also incorporate information about what was likely to occur given all of the facts, and the observer's own interpretation of those events.

Applying theories of intentionality and foresight to causal chains

As in Studies 1 and 2, Study 6 shows that participants' perception of the unitary nature (or unitisation) of the sequence of events does not reflect their selection of the primary cause of the outcome. Thus, the perceptual links between events in these studies cannot be explained by the properties of equifinality as proposed by Heider (1958). Further, the results do not support the prediction that people would see the sequence of events more as a single integrated sequence when the conditions were matched and as two separate incidents when the conditions were mismatched. Instead, these findings suggest that it is the presence of a motive, rather than the valence of the motive that guides people's judgments of how events are linked in these causal chains.

Sloman et al.'s (2012) causal model is well placed to explain the valence effects and the compensatory relations between consecutive events in these causal chains. The findings in Study 6 support the model's bottom-up prediction that people discount intentional actions when the

outcome provides a normative reason to act but not when it opposes action. Indeed, supporting the model's claim that one cause is discounted in favour of another only when the outcome provides a reason to act, the valence of the agent's motive had no effect on causal ratings when there was no normative reason to change the positive outcome. In contrast, participants discounted the agent's causal effect in favour of proximal physical events when there was good reason to change the negative outcome but the positive action adhered to normative standards. In addition, they gave lower judgments of intentionality and foresight when the proximal physical events were considered the better causal explanation of the outcome.

These findings also support the model's top-down hypothesis that people's evaluative reaction to the valence of the agent's intent and the outcome prompts compensatory relations between intentional actions and other possible causes. As shown in these results, participants' causal ratings for proximal events were amplified when the agent's intent violated normative action or unintended harm eventuated, whereas when the intended harm eventuated, the agent's causal effect was amplified while the proximal physical event was diminished.

The new baseline unstated motive condition tested Sloman et al.'s (2012) claim about the conditions necessary for discounting of one cause in favour of another to occur, and the model's top-down inference mechanism. As hypothesized, when things ended well there was no evidence of discounting of intentional actions or compensatory evaluations of distal and proximal events. Participant's inferred intentionality when the intended and foreseen outcome eventuated, and saw both events in the chain as equally good causes of the outcome. However, when harm eventuated, evidence of compensatory relations and discounting of events did emerge, but the specific response patterns were not exactly as predicted.

Indeed, these results do not support the prediction that participants would discount the proximal cause in favour of intentional actions, and that causal ratings would be amplified in response to the norm violations. Instead, the patterns were comparable to those of positive actions that went awry; participants' discounted the agent's causal effect in favour of the proximal physical event. This effect cannot be explained by a desire to blame or that the outcome was morally bad because the agent's action was not necessarily bad. Indeed, these judgments suggest that they assume, that although foreseeable, the agent's motive was most likely principled, and that they would not have intended, or intentionally caused harm. Interestingly, this baseline judgment that others generally mean well and conform to positive actions suggests that Sloman et

al.'s bottom-up inference mechanism primarily shapes people's judgments of intentionality and causality, and that the top-down inference mechanism amplifies ratings for norm violations.

In sum, the key results from this study show that foresight and intentionality play an important role in lay people's attributional reasoning. However, these findings show that it is not intentionality or foresight *per se* that are the critical criteria in judgments of cause. Instead, it is the match between the valence of the agent's motive and the realised outcome that shapes people's perceptions of another's action and the foreseeability of the outcome. These results suggest that this 'mismatch effect' not only shapes people's attributions regarding intentional actions, but also their perception of the causal relations between consecutive events.

Chapter 6: General Discussion

This thesis has examined the effects of the positive or negative valence of the agent's motive and the outcome on lay people's attributions for events in causal chains. The results from these studies support the importance that psychological and legal theorists place on their claim that human agency, intentionality, abnormality, and foresight are core factors in attributional reasoning (e.g., Hart & Honoré, 1985; Heider, 1958; Kelley, 1967; Malle & Knobe, 1997; Weiner, 1995). The standout finding in this thesis is that the valence of the agent's motive and the outcome determines the relative weightings people place on these factors. Peoples' causal judgments reflect both the valence of the agent's motive and the outcome, whereas their judgments of blame primarily reflect the valence of the agent's motive.

Review of Findings

The studies in this thesis integrate two lines of research that have previously been distinct. First, they extend research on attributions about causal chains with negative actions and outcomes to include chains with positive actions and outcomes. Second, they extend research on judgments about a single event in the chain to judgments about two consecutive events in the causal chain. The first novel finding is the clear difference between chains with positive and negative outcomes. With negative outcomes, the type of distal event affects causal judgments about both events in the chain, in that positive and negative actions are judged differently to each other and to physical causes. In contrast, when chains lead to positive outcomes, these differences seen with negative outcomes disappear; the type of distal event or the valence of the agent's motive makes little difference to causal judgments of either the distal or proximal event.

A second novel finding is that causal ratings of distal actions are complemented by the reverse ratings for proximal events, despite the fact that the proximal event is always a physical event that is held constant across the various conditions. These shifts in judgments of the proximal events are not due to any difference in the cause itself, which is unchanged; so the shift is due to changes in the valence of the distal action. The third novel finding is that the morality of the agent's action shapes peoples moral judgments of blame, regardless of the agent's causal effect or the valence of the outcome. Thus, people's cause and blame judgments follow different patterns depending on whether the outcome is good or bad and the agent's motive. These patterns were first shown in Study 1. Studies 2-5 provided a stronger test of the finding reported in Study 1. The main findings and relevance of these studies are summarised in the next few pages.

Study 2 investigated whether the measure of blame used in Study 1 was inappropriate for moral evaluations of positive actions and physical events. Research shows that the construct of blame presupposes a harmful outcome, and carries emotional connotations as well as assessments of liability (Tetlock et al., 2007; Weiner, 1995). Study 2 addressed this issue by replacing the measure of blame used in Study 1 with Alicke et al.'s (2011) praise/blame measure. The findings demonstrated that the effects of the valence of the agent's motive and outcome on judgments of cause and blame shown in Study 1 are robust and cannot be explained by the measures used.

Study 3 addressed two questions regarding the structural equivalence of the outcome and motive conditions in Study 1. First, did the valence effects for outcome in causal judgments reflect differences in the severity, abnormality or salience of the scenarios outcome (Kahneman & Miller, 1986; Kahneman & Tversky, 1982b; Lombrozo, 2010; McClure et al., 1991; Tetlock et al., 2010)? Second, did the valence effects for motive in causal judgments reflect differences in the event's efficacy to cause the outcome (Lombrozo, 2010; Spellman, 1997)? For example, could the effects in the fire scenario be explained by participants seeing arsonists as more competent than Conservation officers at starting fires? To test these possibilities, I used Mandel's (2003, Expt. 2) probability measure as adapted by McClure et al.'s (2007, Expt. 4). The results in Study 3 refute these alternative explanations of the findings in Study 1.

Two pilot studies for Study 4 discounted the possibility that participants' responses reflected differences in the type of motivation (e.g., instrumental vs. situational motivation), or the types of physical events (e.g., mechanical faults rather than natural events) used as exemplars of the physical event categories (Hitchcock & Knobe, 2009; Lombrozo, 2010; Morris et al., 1999; Reeder et al., 2002).

Study 4 examined whether the abnormality of the events, as proposed by legal theorists (Hart & Honoré, 1985; Moore, 2009), could explain the effects of valence on causal attributions. The distal and proximal event conditions were extended to include normal and abnormal physical events, so that the abnormal physical event was the same type as the normal physical event but was designed to meet the legal criteria necessary to be seen as more causal than intentional actions (Moore, 2009, p. 153). The results demonstrated that lay people do discriminate between normal (e.g., strong wind) and abnormal physical events (e.g., a tornado) as proposed by the legal theorists, but the abnormality of the event had little influence on judgments of cause. This finding suggests that abnormality as a determinant of causality is secondary to that of motive and outcome information.

Study 5 examined whether the facilitatory framing of the events (e.g., Mandel & Vartanian, 2009; Spellman et al., 2000), or participants' biases and evaluative reactions to the scenario (e.g., Baron & Hershey, 1988; Enzle & Hawkins, 1992) accounted for the findings reported thus far. I used the probability measure from Study 3 and revised the action statement to describe the agent in neutral terms and in a role that does not violate normative action (e.g., 'a man' rather than 'an arsonist'). The results showed that the same actions were seen as facilitating the outcome when the anticipated outcome eventuated, but as inhibiting the outcome when things went awry. This suggests that participants' perception of each event's causal contribution to the outcome depended more on the congruence between the valence of the agent's motive and the outcome, than the facilitative nature of the event *per se*.

Studies 4 and 5 also examined whether the effects of human agency, motives, and outcomes on judgments of cause, responsibility and praise/blame in Study 3 extend to different situations. Thus, several new scenarios were adapted from earlier research (e.g., Alicke et al., 2011; Hilton et al., 2010; Hitchcock & Knobe, 2009) and added in Studies 4 and 5. The results from both studies revealed effect sizes and interactions between motive and outcome that were similar to those reported in Studies 1-3, providing evidence that these patterns generalise across differing contexts.

In Study 6, I drew on research on the folk concept of intentionality (Malle & Knobe, 1997), hindsight and actor-observer biases (e.g., Malle et al., 2007; Pezzo, 2011) to investigate whether the novel patterns of attributions could be explained by an extension to Sloman et al.'s (2012) causal model. I tested several hypotheses by adding measures of intent (belief and desire), intentionality, and three measures of foresight. Several key findings emerged. First, is that the same motive-outcome asymmetries in judgments of cause reported in Studies 1 – 5 emerged for judgments of intent, intentionality and the three differing perspectives of foresight. The second key finding in Study 6 is that judgment of cause and blame not only reflect the agent's motive and the outcome, but also reflect the observer's perspective: Judgments of cause were influenced by the participant's perception of whether the outcome was likely to eventuate, in addition to the valence of the outcome and the morality of the actions. In contrast, judgments about praise/blame were principally influenced by the morality of the action, whether participants considered the agent should have foreseen the outcome then played a secondary role. The final key finding is that it is not intentionality *per se* that is the critical determinant of people's judgments of cause and blame in these causal chains. Instead, it is the valence of the agent's motive and the outcome

that shapes people's attributional decisions about the relations between independent events in these chains.

Two Patterns of Judgments: A Mismatch Effect and a Motive Effect

In Study 1, the results suggested that the valence of the outcome was the primary determinant of attributions of cause and blame. As the conditions became more structurally equivalent across Studies 2-6, the attributions for positive and negative outcomes also became more equivalent, whereas the distinctions between the judgments of cause and blame became more pronounced. Importantly, two distinct patterns of interactions emerged suggesting that two very different processes underpin lay attributions: a *mismatch effect*, and a *motive effect*. These findings suggest that whether people follow a mismatch effect or a motive effect depends on the attribution: what was the *cause* of the outcome, or who is *culpable* for the outcome?

Pattern 1: A Mismatch Effect.

Across these six studies, participants' attributions regarding causality and the core factors of intent, intentionality, and foresight show the same pattern that I have named the '*mismatch effect*'. Specifically, these judgments reflect the congruence between the valence of the agent's motive and the outcome. When the valence of the agent's motive matched the realised outcome, (i.e., positive motive preceded a positive outcome; or negative motive preceded a negative outcome), participants judged the outcome was intended and discounted the causal efficacy of proximal physical events in favour of distal actions. The reverse pattern emerged when the outcome went awry and there was a 'mismatch' between motive and outcome (i.e., positive motive preceded a negative outcome; or negative motive preceded a positive outcome). Here, participants judged the outcome was unintended and discounted the causal efficacy of distal actions in favour of proximal physical events. Thus, motives and outcomes are interdependent features that have a joint effect on how people judge the primary cause and ascribe responsibility in these chains of events.

Pattern 2: A Motive Effect.

Attributions of culpability (i.e., judgments of blame and punitiveness), have a different pattern to cause attributions, and follow what I term the '*motive effect*'. Here, motive and outcome information have independent effects on attributions. These results show that the moral intent of actions is the primary determinant of judgments of culpability: Bad actions are blameworthy and deserving of punishment, whereas good actions receive at best a modicum of praise and reward. The valence of the outcome plays a secondary role and amplifies ratings of all

causes: all actions and physical events were rated higher in chains with negative than positive outcomes.

In the following sections, I discuss the relations between the constructs of cause, responsibility and blame. I then discuss how well the psychological and legal theories discussed in Chapters 2 and 5 account for the results. Finally, I consider the implications and limitations of this work, together with avenues for future research.

The Relations between Constructs of Cause, Responsibility and Blame

The results from these studies add to our understanding of the relations between the constructs of cause, responsibility, and blame. Importantly, these studies provide evidence that judgments of cause, responsibility, and blame are interrelated but distinct concepts that are not governed by a single reasoning process. Instead, this research shows that these judgments systematically vary in the weightings that people apply information about intent, intentionality, foresight, motive, and outcome. Indeed, these findings support a growing body of work suggesting that lay reasoning does not follow a single pattern, but multiple distinct patterns (Cushman, 2008; Lombrozo, 2010; Mandel, 2003).

The specific patterns of weightings make sense when we consider that judgments of cause reflect the relation between an event and an effect. In contrast, judgments of responsibility reflect a non-emotive assessment of accountability which does not presuppose judgments of liability for harmful outcomes, whereas judgments of blame reflect people's assessment of the morality of another's action (Hart & Honoré, 1985). Interestingly, the finding that judgments of blame were also influenced by participants' perception of what the agent should have expected supports the notion that reasonable foresight is an important factor in legal debate, particularly in cases of negligence. Furthermore, it suggests that lay people see blame not as a one-dimensional construct, but as a multidimensional construct that differs along three continua: causation, intentionality and foresight¹⁷ (Hart & Honoré, 1985). For example, in these studies when the intended harm eventuated, legal theory suggests that these judgments of blame should be described as 'direct' blame,' as action was performed with intent and the outcome foreseen. In contrast, when the intended harm failed to eventuate, legal theory suggests that these judgments of blame should be described as 'indirect' blame,' as the blame attribution was made in the absence of intentionality.

¹⁷ Legal theory distinguishes between three aspects of blame that differ along the dimensions of intentionality and foresight. See Chapter 1, p. 4.

Implications for Psychological and Legal Theories of Causation

This thesis reveals a more subtle and complex picture than that proposed by classic attribution models (e.g., Heider, 1958; Kelley, 1967; Weiner, 1985), or more recent covariation (e.g., Spellman, 1997), and social functionalist (e.g., Alicke, 2000; Tetlock, 2002) approaches. The findings show that people's judgments of causal responsibility and culpability differ to those proposed by legal theorists (e.g., Hart & Honoré, 1985; Moore, 2009).

Human agency: The critical criterion in attributions.

As noted in Chapter 1, human agency is a central factor in attributional and legal theories (e.g., Alicke, 2000; Hart & Honoré, 1985; Heider, 1958; Spellman, 1997; Tetlock, 2002). This prediction was supported when the scenario described chains with negative actions and negative outcomes. For example, a man deliberately starting a fire was seen as the primary cause of a forest fire rather than a lightning strike or a tornado when the intended harm eventuates. However, the results do not support this prediction when the action was performed with positive intent or the intended outcome did not eventuate. For example, when the man's action was motivated by a desire to save a town, it was judged less causal than a lightning strike and considerably less causal than the subsequent tornados when the town was destroyed. Indeed, the finding that subsequent physical event were seen as contributing equally, or were judged more causal than actions when unintended harm eventuates, challenges the core claim of the relevant attributional and legal theories dealing with voluntary actions.

These results imply that attributional and legal theorists were wrong to assume that human agency would be the overarching criterion in lay people's theories of causality. However, examining people's attributions in chains comprising only of physical events in Studies 1-5 provided a comparative base of data that elucidated the role of human agency. In the baseline condition of two consecutive physical events, the finding that distal and proximal events were rated equally causal for both positive and negative outcomes replicates earlier studies with negative outcomes (McClure et al., 2007), and extends this finding to positive outcomes. However, the physical event's proximity to the outcome, abnormality, or its superior probabilistic contribution had little effect on judgments of cause. These findings challenge covariation and legal models that predict that in chains of two similar events, people see the more recent event as having the greater causal effect on the outcome (e.g., Hart & Honoré, 1985; Spellman, 1997). From a social functionalist perspective (e.g., Tetlock, 2002), this finding suggests that both physical events are considered equally viable targets for intervention.

Notably however, there was an absence of abnormality, proximity, or outcome effects for physical events when the effects of voluntariness on people's causal reasoning are controlled. This finding makes an important clarification: abnormality, proximity, and outcomes are core determinants in attributions, but only when voluntary actions are included in the causal chain. In the absence of voluntary actions, lay attributions appear to reflect the premise of equality. This finding suggests that psychological and legal theorists were right to emphasize the 'special status' of human agency in attributional decisions (e.g., Hart & Honoré, 1985; Heider, 1958; Kelley, 1967; Weiner, 1985), but mistaken in their understanding of the relations between the determinants of causality and the influence of the valence of the outcome on lay attributions.

Attributional Theories of Perceived Causation.

As discussed above, the results from these studies are consistent with previous research that have identified information about intentionality, foresight, abnormality, proximity, control, and the valence of the outcome as core determinants in attributions, but only when the scenario describes negative actions in chains with negative outcomes. Notably however, when the valence of the motive and outcome conditions was incongruent, as when chains lead to unanticipated outcomes, the 'mismatch effect' emerged for judgments of cause and the components of intentionality, whereas the 'motive effect' emerged for judgments of culpability. These findings suggest that people's evaluation of each core determinant is shaped by the valence of the agent's motive and the outcome.

As described in Chapter 1, two theoretical approaches have dominated how people conceptualise and explain their own and others' action: the covariational approach stemming from Kelley's (1967) model of attribution (e.g., Cheng & Novick, 1991; Spellman, 1997), and the functionalist approach founded on Heider's (1958) theory of causation (e.g., Alicke, 2000; Knobe & Malle, 1997; Tetlock, 2002). More recently, researchers and theorists have drawn on Heider's claim that people see intentional actions and unintentional actions in fundamentally different ways to explain people's attributions (e.g., Malle & Knobe, 1997; Sloman et al., 2012). The question of how well these key theories explain these findings is examined next.

The covariational approach to lay attributions.

The finding that measures of probability did not explain participants' causal attributions in these studies replicates earlier research on opportunity chains with negative outcomes (Hilton et al., 2005; Lagnado & Channon, 2008; McClure et al., 2007), and is consistent with earlier challenges to probabilistic approaches (e.g., Spellman, 1997). Indeed, intentional actions were

seen more no more efficacious in producing the outcome than physical events when the valence of the agent's motive was controlled. Furthermore, although judgments of the extent that each event raised (or lowered) the outcome's probability influenced participants' judgments of cause, the event that raised the outcome's probability the most was not selected (or 'credited' in Spellman's terms) as the cause. This implies that people do not perform a naïve type of hierarchical multiple regression (Hilton, 1988; Spellman, 1997), and limit their causal reasoning to changes in the perceived necessity or sufficiency of events to determine the primary cause.

It is interesting to note however, that statistically, the conditionalised sufficiency of distal actions¹⁸ described in Spellman's crediting causality model (1997), is the same as the measure of objective foresight which was the primary predictor of causal attributions in Study 6. This implies that Spellman was right to suggest that intentional actions are judged more causal than physical events because they apply greater 'effective personal force' (Heider, 1958, p. 102) thus covary more strongly with the outcome. But, she was mistaken to assume that this reasoning explains people's preferences for intentional actions in chains comprising two independent events. Rather, people may perceive these probability judgments as sufficiently close in definition to represent a proxy measure of objective foreseeability (Fischhoff, 1975).

The functionalist approach to lay attributions.

Alicke's (2000) culpable control model and Tetlock's (2002) social functionalist approach have been useful in understanding lay attributions for actions that have brought about harm. Both models predicted the patterns of attributions of cause and culpability shown in these studies when the negative actions lead to harmful outcomes. Yet the effects of valence of outcome and motive on judgments of cause and punitiveness in these causal chains cannot be explained by Tetlock's social functionalist approach.

These findings support the culpable control model's (2000) proposal that as people look to confirm their 'initial blame hypothesis', they consider how much control the agent exerted over the outcome. How close the agent comes to achieving their goal, and the degree of contextual ambiguity within the scenario, determines how much blame is ascribed: attributions were amplified with the successful realisation of harm but attenuated when the intended harm failed to eventuate. When there was contextual ambiguity as in the mismatched conditions, they discounted the agent's causal control in favour of alternative events. This supports Alicke's

¹⁸ The conditionalised sufficiency is the sufficiency of the target event minus the sufficiency of all prior events to bring about the outcome (see Chapter 1, p. 6, and Chapter 2, p. 2, for a discussion).

(2000) claim that people's assessment of whether the agent acted intentionally and foresaw the processes critically shapes their causal attributions for all of the events in the causal chain. The findings also support the model's assumption that there is a compensatory relation between people's spontaneous evaluations and control evidence.

However, the findings challenge two premises of Alicke's culpable control model (2000). First, the model posits that it is people's ability to control others' action that leads them to identify human actions as more influential causes than environmental factors. Second, although the model focuses on blame, Alicke suggests that the same reasoning applies to attributions of praise. Thus, people's positive evaluative reactions and desire to praise should amplify their judgments of causal control. The present studies show that participants did not distinguish between positive actions and physical events with positive outcomes. Instead, participants rated both events as morally neutral and equally causal, suggesting that they were less inclined to praise good actions and acknowledge their causal effect than Alicke proposed.

Together, these results support the functionalist approach that claims that people are motivated to blame (Alicke, 2000) and think like 'intuitive prosecutors' (Tetlock, 2002, p. 452) to identify and sanction wrongdoers. Furthermore, the extent that people exercise social control appears to extend beyond those who cause harm, but also to such people that could do. However, neither Tetlock's (2002) nor Alicke's (2000) models are able to account for the influence that positive intent has on people's judgments of retribution and culpability. Nevertheless, these findings do support the idea that people are also motivated to be fair (Tetlock et al., 2007). Indeed, these results suggest that in the absence of wrongdoing, people look at alternative explanations to reflect the magnitude of the harmful event.

The concept of intentionality.

In Study 6, I expanded the theoretical approach in this thesis to investigate whether an extension of Sloman et al.'s (2012) causal explanation of the side-effect effect¹⁹ better accounted for the novel effects of valence on attributions. Sloman et al.'s (2012) causal explanation of the side-effect effect provided a useful framework for understanding these effects for several reasons. Firstly, Sloman et al. sought to explain the asymmetries in judgments of intentionality of foreseen side-effects by combining Malle and Knobe's (1997) concept of intentionality with several accounts that cite normative violations, informativeness, evaluative reactions, and discounting as

¹⁹ The side effect-effect refers to the tendency for people to see negative side-effects as intentionally brought about but positive side-effects as unintentionally brought about (Knobe, 2003).

explanations for the effect (e.g., Machery, 2008; Nanay, 2010; Sripada, 2012; Uttich & Lombrozo, 2010). Secondly, Sloman et al.'s (2012) model proposes that there are two inferential mechanisms that shape people's judgments of intentionality and cause: a bottom-up mechanism and a top-down mechanism. Specifically, the bottom-up mechanism implies that people evaluate information from the effect to the cause and involves discounting of intent and causal efficacy, whereas the top-down mechanism suggests that people evaluate information from the agent's character to the action and amplify attributions. These premises are consistent with the findings here and with other evidence that attributional reasoning is not confined to a single process (e.g., Cushman, 2008; Mandel, 2003; Lombrozo, 2010).

Finally, Sloman et al.'s (2012) model also proposes that people's assessment of the agent's intent depends not only on what the agent's epistemic state was at the time the action was performed, but also on whether that state made a difference, and whether their judgments would vary were that to change. Thus, given the backward looking nature of attributional decisions, people's attributions are sensitive to factors such as perspective, world knowledge, biases, and the valence of the conditions that earlier researchers linked to causal reasoning (e.g., Alicke, 2000; Kashima et al., 1998; Tetlock, 2002).

Two key findings in Study 6 suggest that extending Sloman et al.'s (2012) causal model to chains with multiple events provides the most parsimonious account of the findings in this thesis. The first key finding is that participants discounted the distal action's causal effect in favour of subsequent events with negative but not positive outcomes. The second key finding is that when chains lead to positive outcomes, participants discounted intent (belief and desire), intentionality, and foresight when the agent's action violated social norms, but amplified these measures when the same action adhered to social norms. Taken together, these findings suggest that people take a holistic approach and 'intuitively balance the causal equation' by discounting or amplifying the causal effect of each event in the causal chain when there is a moral or ethical reason for action, but not when it opposes action. Finally, the findings suggest that people interpret the agent's causal role and the component factors of intentionality in a manner that is consistent with the outcome of the scenario rather than based on the facts about the action described in the action statement (Kashima et al., 1998).

The Importance and Implications of Motive Statements in Lay Reasoning

Extending this research to chains that manipulated the presence of motive information in Study 6 has shown that although the valence of the outcome is important, the agent's motive

plays the critical role in shaping people's attributions of cause and culpability in sequences of multiple events. In addition, the agent's motive rather than voluntariness *per se* determines when physical events are relegated to the background. When motive information was absent, the presence of intentional actions or the valence of the outcome makes little difference; the proximal physical event is judged to have caused the outcome. This finding replicates earlier research on valence effects (Ybarra, 2002) that suggests that people's baseline assumption is that the agent acted with positive intent. However, when the motive is stated, people amplify or discount the agent's causal effect consistent with the realised, rather than the anticipated outcome. The role of subsequent physical events then becomes a compensatory means to balance the causal equation. It is as if the absence of a motive results in people making more impartial evaluations of each event's causal contribution to the outcome. In contrast, the presence of a motive appears to prompt people to focus squarely on the voluntary action and make compensatory rather than event-specific judgments about the other cause.

The findings in this thesis have important ramifications when people are asked to decide questions of causal responsibility and culpability in courts of law. These results show that in cases where the defendant's harmful intent was realised with a negative outcome, the defendant's causal responsibility for the harm is reflected in people's judgments of legal liability (i.e., blame and punitiveness). Thus lay attributions concur with legal principles (Moore, 2009). However, when intentional actions have unintended consequences, as in cases of manslaughter or inchoate crime, these studies suggest that lay people do not follow the legal claim that judgments of culpability are apportioned 'by degrees of causal contribution' (Moore, 2009, p. 118). For example, when unintended harm eventuated, physical events that followed positive actions mitigated judgments of the agent's culpability. In contrast, when the expected harm failed to eventuate, the unintended positive outcome did little to mitigate judgments of culpability. These findings suggest that people's 'retributive wrath' (Tetlock et al., 2007, p. 196), is provoked in response to norm violations and that their judgments of culpability are somewhat harsher than the legal claim that successful crimes are worse than inchoate crimes by a factor of 2 (Moore, 2009, p. 21).

Broader Methodological Issues and Future Directions

Extending attributional research in these studies to chains of multiple target events addresses the concern that constraining participants' judgments to single target events limits our ability to generalise findings to the 'real' world. Furthermore, the effects of valence on

participants' attributions in these studies suggests that earlier attributional research with only negative valence conditions did not capture the complexity of people's attributional decisions (e.g., Cushman, 2008; Hilton et al., 2010; McClure et al., 2007). These studies used the scenario paradigm which included two consecutive, target events. Of course, different results might be attained with different methods. But the scenario paradigm is still widely used in psychology (e.g., Callan, Sutton, & Dovele, 2010; Mandel, 2013) and law (e.g., Charron & Lowe, 2008; Nadler & McDonnell, 2012), and is based on real world situations that make up the scenarios. In addition, given the consistent valence effects found across these studies, suggests that participants engaged with the text when making attributions. However, examining attributions in complex causal structures increases the risk of alternative explanations of the effects. As noted above, several revisions were made across Studies 2-5 to ensure structurally equivalent scenarios. The results show that the reported valence effects cannot be explained by differences between scenarios.

Despite the revisions made to ensure structurally equivalent conditions, one criticism that could be made concerns the use of the scenario paradigm where participants knew that the outcome in question had in fact occurred. It is possible that their judgments were susceptible to hindsight bias and effects of valence were overestimated (Fischhoff, 1975). Replicating the hindsight paradigm where outcome information is withheld in the foresight condition, a no-outcome information condition is viable for evaluating the target event for moral judgments such as praise/blame, or judgments of intent and subjective foresight. But withholding outcome information is problematic when people are judging the relation between an event and an effect, as for judgments of causality, intentionality, or objective and reasonable foresight. Specifically, can people make these judgments if there is no outcome?

Research has shown that people are endowed with 'means-end' schema and envisage outcomes in the absence of outcome information (Alicke et al., 2008; Epstude & Roese, 2008; Kahneman & Miller, 1986; Schank & Abelson, 1977). Alicke et al.'s (2008, Expt. 1) address this issue with their inclusion of a no-outcome condition for judgments of praise/blame but not cause. The authors reported that the responses and effect sizes for judgments of praise/blame when the agent's action either conformed or violated norms in the no-outcome condition closely replicated those in their matched conditions. For example, when the physician followed hospital procedures and the patient survived, his decision to act was considered praiseworthy (norm conforming), but blameworthy when he ignored procedures and the patient died (norm violating).

It is also possible that the valence of the agent's motive or outcome would moderate people's perceptions of the functional utility of the event. For example, the event people see as the primary cause might change as the counterfactual focus moves from prevention to promotion with positive outcomes (Alicke et al., 2008; Roese, Hur, & Pennington, 1999). Hence, it is likely that judgments would show a different pattern if participants are asked to judge an agent's intent (belief and desires) and make prospective estimates of cause and intentionality before the outcome is known.

Therefore, given that parallels may be drawn between Alicke et al.'s (2008) findings and the motive effects reported in these studies, I considered it unlikely that a no-outcome condition would function well in these causal chains for several reasons. First, while addressing the question of hindsight bias, it is likely that a no-outcome condition would nevertheless be susceptible to counterfactual biases, including heuristics and expectancy biases. Second, participants may become frustrated with the lack of direction given limited contextual information, or with a scenario that makes no sense to them. Finally, an additional condition would add further complexity to an already complex experimental design.

From a broader perspective, these findings add to the growing body of research that suggests that causal judgments serve several functions, each of which employs a differing reasoning process (e.g., Cushman, 2008; Lombrozo, 2010; Mandel, 2003). For example, Mandel (2003) showed that lay people employed causal, counterfactual, or covariational reasoning depending on the functional objective of the causal question: for the purpose of retribution, future prevention, or promotion. Mandel also showed that the relative weightings people placed on the causal information differed depending on these functional objectives, which in turn lead to predictable dissociations in the events selected as the 'cause' in chains with negative outcomes. Therefore, it might be that people weigh the components of intentionality and foresight depending on the functional objective of the causal judgment. Returning to the fire, for example; the tornado may be identified as the cause of dozens of homes being destroyed by fire as a mitigating factor in cases of negligence. Alternatively, people might identify the man as the cause of the outcome for the purpose of future interventions, if they consider he failed to accurately assess all environmental risks prior to acting. Future experiments could investigate the mismatch effect on people's judgments of cause when the functional objective of the attribution is manipulated. Such designs would help legal, medical and human resource practitioners ensure that interventions are targeted most effectively, or critical events are appropriately identified for

training or recognition purposes. Similarly, business strategists may benefit from analysing a problem from differing focal and reasoning perspectives.

Research has shown that people are sensitive to the temporal position of voluntary actions and physical events in causal chains leading to negative outcomes (Gerstenberg & Lagnado, 2012; Hilton et al., 2010; Lagnado & Channon, 2008; McClure et al., 2007). Taken together, these findings suggest that the effect of temporal order indicates that primary causes that are proximal to the outcome are judged better causal explanations than distal events for reasons other than voluntariness, intentionality, abnormality, or changes in outcome probability. Segura and McCloy (2003) found that participants' counterfactual analysis of the causal event depended not only on the abnormality of the events, but also on the length of the causal chain and the target event's temporal position. Examining the mismatch effect on people's judgments of cause as a function of changes in the temporal order of the causes would elucidate how well this account extends to differing contexts.

Another extension of the current studies would be to combine the valence and intentionality manipulations with longer causal chains. This is important, as foreseeability might interact with temporal position of the events in the chain. For example, Hilton et al. (2005) showed that when distal events had foreseeable effects on the outcome, this moderates judgments regarding subsequent events. A further issue concerns the distinctions between intentional and unintentional actions involving negligence or accidental outcomes. Malle and Knobe (1997) showed that when any one of the five components of intentionality was absent, people see the action as unintentional. Yet, it is possible that people do not see each of these components as equally important criteria of intentionality. For example, when all other factors were held constant, Hilton and McClure (2013) found that judgments systematically varied depending on the agent's epistemic state. Whether the agent knowingly took advantage, or unwittingly profited from the opportunity to cause the effect had a stronger effect on judgments of responsibility than cause depending on the deliberateness of the action. Lagnado and Channon (2008) found that judgments of cause and blame diverged for judgments of foresight depending on the observer's perspective. Therefore, if we assume that intentionality, like causality, is continuous rather than binary, it is reasonable to think that negligent actions and accidental actions would be seen differently as people assess the agent's intent, skill, foresight and awareness along each continuum and weigh each factor according to a hierarchy of importance. This has important

implications regarding the emphasis placed on the differing aspects of evidence when we are asked to judge another's action, whether in courts of law, or in the workplace.

Finally, one criticism that could be made of this thesis regards the complexity of the studies' experimental design. As noted above, several variables have been examined to explain how people explain and interpret the world around them. Certainly, similar breadth of variables is found in research across social psychology (e.g., Ellis, Schlomer, Tilley, & Butler, 2012), which in itself reflects the complexity of lay reasoning. Yet the challenges this thesis poses for key aspects of attributional and legal theories suggest that their approach is too simplistic and fails to capture key variables. Indeed, by focusing on single target events or comparing single valenced conditions, e.g., negative motives against neutral conditions, the important relations between variables have been obscured and the responses reflect only one part of the story.

These studies are indeed complex, but not because irrelevant or trivial variables have been included. When you strip down to the core findings of judgments of cause and blame and put aside the findings for responsibility, punitiveness, intentionality, and foresight, the patterns are the same. Indeed out of this complexity, two simple effects explain all of the findings: a mismatch effect that explains judgments of cause and intentionality; and a motive effect that explains judgments of culpability.

Conclusion

This thesis clarifies the relations between theoretical and folk concepts of causation and extends links to legal theories. It also bridges the gap between research on the effects of valence on attributions for single target events and attributions in chains of multiple events. In doing so, this thesis shows that each of the factors identified in legal concepts such as *mens rea* are important when considered in isolation. However, it has also shows that it is not the discrete influence of each factor that is the key determinant in causal attributions, but how people reconcile the agent's role in the sequence with what actually eventuated.

Importantly, the findings across six studies provide strong evidence demonstrating how the positive and negative valence of the agent's motive and the outcome influence lay people's attributions in chains of multiple events. The findings also show that people do not make event-specific judgments, but take a more holistic view of the causal chain and allocate cause in a compensatory way.

Finally, this thesis establishes a link between research in social psychology, folk psychology and experimental philosophy, to address several questions regarding the relations between judgments of cause, judgments of blame, and questions of liability. Although these relations have been examined extensively, one significance of this thesis lies in the finding that judgments of cause, responsibility, blame, and punitiveness are not governed by a single reasoning process. Instead, judgments about these the concepts vary by making use of information regarding the components of intentionality, the agent's motive and the outcome in different ways.

In summary, this thesis goes some way to unravelling the anomalies in earlier research regarding the interplay between valence and lay reasoning in attributions. The findings also show lay people's attributions are not always in agreement with theoretical approaches. In closing, this thesis provides a cautionary tale about the pervasive effect of negative information and the importance of the differing perspectives of foresight to ensure that we understand how the scales of justice are intuitively balanced.

References

- Abelson, R. P., & Lalljee, M. (1988). Knowledge structures and causal explanations. . In D. J. Hilton (Ed.), *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality*. Brighton: Harvester.
- Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 63(3), 368-378.
- Alicke, M. D. (1994). Evidential and extra-evidential evaluations of social conduct. *Journal of Social Behavior & Personality*, 9(4), 591-615.
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126(4), 556-574.
- Alicke, M. D. (2008). Blaming Badly. *Journal of Cognition & Culture*, 8(1/2), 179-186. doi: 10.1163/156770908x289279
- Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin*, 34(10), 1371-1381. doi: DOI: 10.1177/0146167208321594
- Alicke, M. D., Davis, T. L., & Pezzo, M. V. (1994). A posteriori adjustment of a priori decision criteria. *Social Cognition*, 12(4), 281-308.
- Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy*, 108(12), 670.
- Alicke, M. D., Yurak, T. J., & Vredenburg, D. S. (1996). Using personal attitudes to judge others: The roles of outcomes and consensus. *Journal of Research in Personality*, 30(1), 103-119.
- Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54(4), 569-579. doi: <http://dx.doi.org/10.1037/0022-3514.54.4.569>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323-370.
- Bohner, G., Bless, h., Schwarz, N., & Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology*, 18(4), 335-345.
- Brockner, J., Fishman, A. Y., Reb, J., Goldman, B., Spiegel, S., & Garden, C. (2007). Procedural fairness, outcome favorability, and judgments of an authority's responsibility. *Journal of*

- Applied Psychology*, 92(6), 1657-1671. doi: <http://dx.doi.org/10.1037/0021-9010.92.6.1657>
- Callan, M. J., Sutton, R. M., & Dovele, C. (2010). When deserving translates into causing: The effect of cognitive load on immanent justice reasoning. *Journal of Experimental Social Psychology*, 46(1), 25. doi: 10.1016/j.jesp.2010.05.024
- Catellani, P., Alberici, A. I., & Milesi, P. (2004). Counterfactual thinking and stereotypes: The nonconformity effect. *European Journal of Social Psychology*, 34(4), 421-436. doi: 10.1002/ejsp.206
- Charron, K. F., & Lowe, D. J. (2008). An examination of the influence of surprise on judges and jurors' outcome effects. *Critical Perspectives on Accounting*, 19(7), 1020-1033. doi: <http://dx.doi.org/10.1016/j.cpa.2007.06.005>
- Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, 40(1-2), 83-120. doi: [http://dx.doi.org/10.1016/0010-0277\(91\)90047-8](http://dx.doi.org/10.1016/0010-0277(91)90047-8)
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The Problem of Units and the Circumstance for POMP. *Multivariate Behavioral Research*, 34(3), 315-346. doi: 10.1207/S15327906MBR3403_2
- Cole Wright, J., & Bengson, J. (2009). Asymmetries in Judgments of Responsibility and Intentional Action. *Mind & Language*, 24(1), 24-50. doi: 10.1111/j.1468-0017.2008.01352.x
- Creyer, E. H., & Gürhan, Z. (1997). Who's to blame? Counterfactual reasoning and the assignment of blame. *Psychology & Marketing*, 14(3), 209-222. doi: 10.1002/(sici)1520-6793(199705)14:3<209::aid-mar1>3.0.co;2-f
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108(2), 353-380.
- Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, 108(1), 281-289. doi: 10.1016/j.cognition.2008.02.005
- Ellis, B. J., Schlomer, G. L., Tilley, E. H., & Butler, E. A. (2012). Impact of fathers on risky sexual behavior in daughters: A genetically and environmentally controlled sibling study. *Development and Psychopathology*, 24(1), 317-332. doi: <http://dx.doi.org/10.1017/S095457941100085X>

- Enzle, M. E., & Hawkins, W. L. (1992). A priori actor negligence mediates a posteriori outcome effects on moral judgment. *Journal of Experimental Social Psychology*, 28(2), 169-185. doi: [http://dx.doi.org/10.1016/0022-1031\(92\)90037-K](http://dx.doi.org/10.1016/0022-1031(92)90037-K)
- Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, 12(2), 168-192.
- Fincham, F., & Jaspars, J. (1979). Attribution of responsibility to the self and other in children and adults. *Journal of Personality and Social Psychology*, 37(9), 1589-1602.
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288-299. doi: <http://dx.doi.org/10.1037/0096-1523.1.3.288>
- Forgas, J. P. (2008). Affect and cognition. *Perspectives on Psychological Science*, 3(2), 94-101. doi: 10.1111/j.1745-6916.2008.00067.x
- Gerstenberg, T., & Lagnado, D. A. (2012). When contributions make a difference: Explaining order effects in responsibility attribution. *Psychonomic Bulletin and Review*, 19(4), 729-736.
- Giroto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, 78(1-3), 111-133.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.
- Guglielmo, S., & Malle, B. F. (2010b). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin*, 36(12), 1635-1647.
- Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition*, 25(1), 48-63.
- Hart, H. L. A., & Honoré, A. M. (1985). *Causation in the Law* (Second ed.). Oxford: Oxford University Press
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107(3), 311-317.
- Heider, F. (1958). *The psychology of interpersonal relations* (Vol. 326).
- Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, 107(1), 65-81. doi: <http://dx.doi.org/10.1037/0033-2909.107.1.65>
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, 118(2), 248-271. doi: 10.1037/0033-2909.118.2.248

- Hilton, D. J. (Ed.). (1988). *Logic and Causal Attribution*. Brighton Harvester.
- Hilton, D. J., & McClure, J. (2013). *Acting knowingly: Effects of epistemic state on causal attributions, responsibility judgments and blame*.
- Hilton, D. J., McClure, J., & Sutton, R. M. (2010). Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes? *European Journal of Social Psychology*, 40(3), 383-400.
- Hilton, D. J., McClure, J. L., & Slugoski, B. R. (2005). The course of events: Counterfactuals, causal sequences, and explanation. In D. R. Mandel, D. J. Hilton & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 44 - 60). New York, NY, : US: Routledge.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 106(11), 587-632.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136-153.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge university Press.
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143-157. doi: 10.1016/0010-0277(82)90023-3
- Kahneman, D., & Tversky, A. (1982b). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201 - 208). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341-350. doi: 10.1037/0003-066x.39.4.341
- Kashima, Y., McKintyre, A., & Clifford, P. (1998). The Category of the Mind: Folk Psychology of Belief, Desire, and Intention. *Asian Journal of Social Psychology*, 1(3).
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on Motivation* (Vol. 15, pp. 192-238). Lincoln: University of Nebraska Press.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 28, 107 - 128.
- Knobe, J. (2003). Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology*, 16(2), 309.
- Knobe, J. (2010). Person as scientist, person as moralist. . *Behavioral and Brain Sciences*, 33, 315-365. doi: 10.1017/S0140525X10000907
- Kun, A., & Weiner, B. (1973). Necessary versus sufficient causal schemata for success and failure. *Journal of Research in Personality*, 7(3), 197-207.

- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, 108(3), 754-770.
- Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, 61(4), 303-332.
- Machery, E. (2008). The Folk Concept of Intentional Action: Philosophical and Experimental Issues. *Mind & Language*, 23(2), 165-189. doi: 10.1111/j.1468-0017.2007.00336.x
- Malle, B. F. (2006a). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, 132(6), 895-919.
- Malle, B. F. (2006b). Intentionality, Morality, and Their Relationship in Human Judgment. *Journal of Cognition and Culture. Special Issue: Folk conceptions of mind, agency and morality*, 6(1-2), 87-112.
- Malle, B. F. (2011). Time to Give Up the Dogmas of Attribution: An Alternative Theory of Behavior Explanation. In M. O. James & P. Z. Mark (Eds.), *Advances in Experimental Social Psychology* (Vol. Volume 44, pp. 297-352): Academic Press.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33(2), 101-121.
- Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 93(4), 491-514.
- Mallon, R. O. N. (2008). Knobe versus Machery: Testing the Trade-Off Hypothesis. *Mind & Language*, 23(2), 247-255. doi: 10.1111/j.1468-0017.2007.00339.x
- Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual and covariational reasoning. *Journal of Experimental Psychology: General*, 132(3), 419-434. doi: 10.1037/0096-3445.1132.3.419
- Mandel, D. R. (2013). Do Framing Effects Reveal Irrational Choice? *Journal of Experimental Psychology: General*. doi: <http://dx.doi.org/10.1037/a0034207>
- Mandel, D. R., & Vartanian, O. (2009). Weighting of contingency information in causal judgement: Evidence of hypothesis dependence and use of a positive-test strategy. *The Quarterly Journal of Experimental Psychology*, 62(12), 2388-2408. doi: 10.1080/17470210902794148
- Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the Robustness of Outcome Bias: No Constraint by Prior Culpability. *Basic and Applied Social Psychology*, 26(2-3), 131-146.

- McClure, J. (1984). On necessity and commonsense: A discussion of central axioms in new approaches to lay explanation. *European Journal of Social Psychology*, 14(2), 123-149. doi: 10.1002/ejsp.2420140202
- McClure, J. (1998). Discounting causes of behavior: Are two reasons better than one? *Journal of Personality and Social Psychology*, 74(1), 7-20.
- McClure, J., & Hilton, D. (1997). For you can't always get what you want: When preconditions are better explanations than goals. *British Journal of Social Psychology*, 36(2), 223-240.
- McClure, J., & Hilton, D. J. (1998). Are goals or preconditions better explanations? It depends on the question. *European Journal of Social Psychology*, 28(6), 897-911.
- 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. Special Issue: Familiarity impacts person perception*, 37(5), 879-901.
- McClure, J., Lalljee, M., & Jaspars, J. (1991). Explanations of extreme and moderate events. *Journal of Research in Personality*, 25(2), 146-166.
- Menec, V. H., & Weiner, B. (2000). Observers' Reactions to Genetic Testing: The Role of Hindsight Bias and Judgments of Responsibility¹. *Journal of Applied Social Psychology*, 30(8), 1670-1690. doi: 10.1111/j.1559-1816.2000.tb02461.x
- Moore, M. S. (2009). *Causation and Responsibility*. Oxford: Oxford University Press.
- Morris, M. W., Moore, P. C., & Sim, D. L. H. (1999). Choosing remedies after accidents: Counterfactual thoughts and the focus on fixing "human error." *Psychonomic Bulletin & Review*, 6(4), 579-585.
- 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(2), 139-162. doi: 10.1006/jesp.1995.1007
- Nadler, J., & McDonnell, M. H. (2012). Moral character, motive, and the psychology of blame. *Cornell Law Review*, 97(2), 255-304.
- Nanay, B. (2010). Morality or Modality? What Does the Attribution of Intentionality Depend On? *Canadian Journal of Philosophy*, 40(1), 25-39, 152.
- Naquin, C. E., & Kurtzberg, T. R. (2004). Human reactions to technological failure: How accidents rooted in technology vs. human error influence judgments of organizational accountability. *Organizational Behavior and Human Decision Processes*, 93(2), 129-141.

- Nelson-le Gall, S. A. (1985). Motive–outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology*, 21(2), 332-337. doi: <http://dx.doi.org/10.1037/0012-1649.21.2.332>
- Ohtsubo, Y. (2007). Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect. *Japanese Psychological Research*, 49(2), 100-110.
- Pezzo, M. V. (2011). Hindsight Bias: A Primer for Motivational Researchers. *Social and Personality Psychology Compass*, 5(9), 665-678. doi: 10.1111/j.1751-9004.2011.00381.x
- Reeder, G. D., Kumar, S., Hesson-McInnis, M., & Trafimow, D. (2002). Inferences about the morality of an aggressor: The role of perceived motive. *Journal of Personality and Social Psychology*, 83(4), 789-803. doi: 10.1037/0022-3514.83.4.789
- Reeder, G. D., & Trafimow, D. (2005). *Attributing Motives to Other People* Malle, Bertram F.; Hodges, Sara D (pp. (2005). Other minds: How humans bridge the divide between self and others. (pp. 2106-2123). New York, NY, US: Guilford Press. xiii, 2354).
- Robbennolt, J. K. (2000). Outcome Severity and Judgments of “Responsibility”: A Meta-Analytic Review1. *Journal of Applied Social Psychology*, 30(12), 2575-2609. doi: 10.1111/j.1559-1816.2000.tb02451.x
- Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121(1), 133-148. doi: 10.1037/0033-2909.121.1.133
- Roese, N. J., Hur, T., & Pennington, G. L. (1999). Counterfactual thinking and regulatory focus: Implications for action versus inaction and sufficiency versus necessity. *Journal of Personality and Social Psychology*, 77(6), 1109-1120. doi: 10.1037/0022-3514.77.6.1109
- Ryder, J. (1997). *Explaining events: The effect of voluntariness and abnormality on causal attribution*. PSYC 410: Research Topic. Victoria University of Wellington.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.
- Schkade, D. A., & Kilbourne, L. M. (1991). Expectation-outcome consistency and hindsight bias. *Organizational Behavior and Human Decision Processes*, 49(1), 105-123. doi: [http://dx.doi.org/10.1016/0749-5978\(91\)90044-T](http://dx.doi.org/10.1016/0749-5978(91)90044-T)
- Segura, S., & McCloy, R. (2003). Counterfactual thinking in everyday life situations: Temporal order effects and social norms. *Psicológica*, 24(1), 1-15.

- Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility and blameworthiness*. New York: Springer-Verlag.
- Shaw, J. C., Wild, E., & Colquitt, J. A. (2003). To justify or excuse?: A meta-analytic review of the effects of explanations. *Journal of Applied Psychology*, 88(3), 444-458. doi: <http://dx.doi.org/10.1037/0021-9010.88.3.444>
- Sloman, S. A., Fernbach, P. M., & Ewing, S. (2012). A Causal Model of Intentionality Judgment. *Mind and Language*, 27(2), 154-180.
- Spellman, B. A. (1997). Crediting Causality. *Journal of Experimental Psychology: General*, 126(4), 323-348.
- Spellman, B. A., Price, C., & Logan, J. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory & Cognition*, 29(2), 193-208. doi: 10.3758/bf03194913
- Sripada, C. S. (2012). Mental state attributions and the side-effect effect. *Journal of Experimental Social Psychology*, 48(1), 232-238.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, 109(3), 451-471. doi: <http://dx.doi.org/10.1037/0033-295X.109.3.451>
- Tetlock, P. E., Self, W. T., & Singh, R. (2010). The punitiveness paradox: When is external pressure exculpatory - And when a signal just to spread blame? *Journal of Experimental Social Psychology*, 46(2), 388-395.
- Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, S. B., . . . Rescober, P. (2007). People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility. *Journal of Experimental Social Psychology*, 43(2), 195-209.
- Tostain, M., & Lebreuilly, J. (2008). Rational model and justification model in 'outcome bias.'. *European Journal of Social Psychology*, 38(2), 272-279. doi: 10.1002/ejsp.404
- Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116(1), 87-100.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York, NY, US: Guilford Press. (1995) xvi, 301 pp.
- Wells, G. L., Taylor, B. R., & Turtle, J. W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53(3), 421-430. doi: 10.1037/0022-3514.53.3.421

- White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin*, Vol., 108(1), 3-18.
- Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100(2), 283-301.
- Ybarra, O. (2002). Naive causal understanding of valenced behaviors and its implications for social information processing. *Psychological Bulletin*, 128(3), 421-441. doi: <http://dx.doi.org/10.1037/0033-2909.128.3.421>
- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), 8235-8240.
- Zhu, J., & Thagard, P. (2002). Emotion and action. *Philosophical Psychology*, 15(1), 19-36. doi: 10.1080/09515080120109397

Appendix A: Scenarios

Studies 1-3

Studies 1-3 used a mixed design with three between-subject variables: distal event (positive action, negative action, or a physical event), outcome (positive or negative) and scenario (fire, train, or factory). The proximal physical event was the same for all conditions in each scenario. The 3 (distal event) x 2 (outcome) design resulted in six versions of each scenario.

Scenario: Forest Fire

1. Positive outcome: Distal positive action

A Conservation Department officer deliberately set fire to an area of forest. His goal was to cause a burn-off to prevent a larger fire from spreading. A strong wind developed which fanned the flames. These events prevented the larger fire from reaching the nearby town, saving dozens of homes.

2. Positive outcome: Distal negative action

An arsonist deliberately set fire to an area of forest. His goal was to cause a larger fire to spread. A strong wind developed which fanned the flames. These events prevented the larger fire from reaching the nearby town, saving dozens of homes.

3. Positive outcome: Distal physical event

A lightning strike set fire to an area of forest. A strong wind developed which fanned the flames. These events prevented the larger fire from reaching the nearby town, saving dozens of homes.

4. Negative outcome: Distal positive action

A Conservation Department officer deliberately set fire to an area of forest. His goal was to cause a burn-off to prevent a larger fire from spreading. A strong wind developed which fanned the flames. These events led to the larger fire reaching the nearby town, destroying dozens of homes.

5. Negative outcome: Distal negative action

An arsonist deliberately set fire to an area of forest. His goal was to cause a larger fire to spread. A strong wind developed which fanned the flames. These events led to the larger fire reaching the nearby town, destroying dozens of homes.

6. Negative outcome: Distal physical event

A lightning strike set fire to an area of forest. A strong wind developed which fanned the flames. These events led to the larger fire reaching the nearby town, destroying dozens of homes.

Scenario: Train Accident

1. Positive outcome: Distal positive action

A railway worker changed the points at a junction on a railway track. His goal was to guide an oncoming passenger train onto a second track to prevent a train collision. A short time later the passenger train was re-routed onto the second track. A faulty barrier arm did not lower and stop traffic at the next railway crossing. These events prevented the train from colliding with a full school bus that was on the level crossing saving dozens of lives.

2. Positive outcome: Distal negative action

A terrorist changed the points at a junction on a railway track. His goal was to guide an oncoming passenger train onto a second track to cause a train collision. A short time later the passenger train was re-routed onto the second track. A faulty barrier arm did not lower and stop traffic at the next railway crossing. These events prevented the train from colliding with a full school bus that was on the level crossing saving dozens of lives.

3. Positive outcome: Distal physical event

Rocks from a landslide changed the points at a junction on a railway track. A short time later a passenger train was re-routed onto a second track. A faulty barrier arm did not lower and stop traffic at the next railway crossing. These events prevented the train from colliding with a full school bus that was on the level crossing saving dozens of lives.

4. Negative outcome: Distal positive action

A railway worker changed the points at a junction on a railway track. His goal was to guide an oncoming passenger train onto a second track to prevent a train collision. A short time later the passenger train was re-routed onto the second track. A faulty barrier arm did not lower and stop traffic at the next railway crossing. These events caused the train to collide with a full school bus that was on the level crossing resulting in the loss of dozens of lives.

5. Negative outcome: Distal negative action

A terrorist changed the points at a junction on a railway track. His goal was to guide an oncoming passenger train onto a second track to cause a train collision. A short time later the passenger train was re-routed onto the second track. A faulty barrier arm did not lower and stop

traffic at the next railway crossing. These events caused the train to collide with a full school bus that was on the level crossing resulting in the loss of dozens of lives.

6. Negative outcome: Distal physical event

Rocks from a landslide changed the points at a junction on a railway track. A short time later a passenger train was re-routed onto a second track. A faulty barrier arm did not lower and stop traffic at the next railway crossing. These events caused the train to collide with a full school bus that was on the level crossing resulting in the loss of dozens of lives.

Scenario: Factory

1. Positive outcome: Distal positive action

In a large poultry processing plant, a diligent employee adjusted the thermostat on a conveyor belt. His goal was to ensure the chicken meat met temperature safety standards to eliminate the risk of food poisoning. At the end of the conveyor belt, a faulty automatic control unit reset all of the plant's thermostats. These events prevented a deadly outbreak of food poisoning saving dozens of lives.

2. Positive outcome: Distal negative action

In a large poultry processing plant, a disgruntled employee adjusted the thermostat on a conveyor belt. His goal was to ensure the chicken meat rose above temperature safety standards to increase the risk of food poisoning. At the end of the conveyor belt, a faulty automatic control unit reset all of the plant's thermostats. These events prevented a deadly outbreak of food poisoning saving dozens of lives.

3. Positive outcome: Distal physical event

A power surge in a large poultry processing plant disrupted the temperature faulty control units on a conveyor belt. At the end of the conveyor belt, a faulty automatic control unit reset all of the plant's thermostats. These events prevented a deadly outbreak of food poisoning saving dozens of lives.

4. Negative outcome: Distal positive action

In a large poultry processing plant, a diligent employee adjusted the thermostat on a conveyor belt. His goal was to ensure the chicken meat met temperature safety standards to eliminate the risk of food poisoning. At the end of the conveyor belt, a faulty automatic control unit reset all of the plant's thermostats. These events led to a deadly outbreak of food poisoning resulting in the loss of dozens of lives.

5. Negative Outcome: Distal negative Action

In a large poultry processing plant, a disgruntled employee adjusted the thermostat on a conveyor belt. His goal was to ensure the chicken meat rose above temperature safety standards to increase the risk of food poisoning. At the end of the conveyor belt, a faulty automatic control unit reset all of the plant's thermostats. These events led to a deadly outbreak of food poisoning resulting in the loss of dozens of lives.

6. Negative Outcome: Distal physical Event

A power surge in a large poultry processing plant disrupted the temperature faulty control units on a conveyor belt. At the end of the conveyor belt, a faulty automatic control unit reset all of the plant's thermostats. These events led to a deadly outbreak of food poisoning resulting in the loss of dozens of lives.

Study 4

Two new scenarios were added in Study 4. The distal and proximal physical events had two versions: a normal or an abnormal event. Thus, Study 4 extended the mixed design used in Studies 1-3, to include four between-subject variables: distal event (positive action, negative action, normal, or abnormal), proximal physical event (normal, or abnormal), and outcome (positive or negative), and scenario (fire, train, factory, avalanche, flooding). The 4 (distal event) x 2 (proximal event) x 2 (outcome) design resulted in sixteen versions of each scenario.

Scenario: Avalanche

1. Positive outcome: Distal positive action: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A ski patroller knew he was risking his life, but he wanted to make sure the ski village at the bottom of the ski field was safe. The patroller climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges whose occupants were snowed in. These events prevented a significant loss of life.

2. Positive outcome: Distal negative action: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A rebel wanted to avenge his brother's death by destroying the ski village at the bottom of the ski field. The rebel climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges. These events prevented a significant loss of life.

3. Positive outcome: Distal normal event: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During a storm, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the

avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges whose occupants were snowed in. These events prevented a significant loss of life.

4. Positive outcome: Distal abnormal event: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During one of the worst storms ever recorded, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges whose occupants were snowed in. These events prevented a significant loss of life.

5. Positive outcome: Positive action: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A ski patroller knew he was risking his life, but he wanted to make sure the ski village at the bottom of the ski field was safe. The patroller climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges whose occupants were snowed in. These events prevented a significant loss of life.

6. Positive outcome: Negative action: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A rebel wanted to avenge his brother's death by destroying the ski village at the bottom of the ski field. The rebel climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges whose occupants were snowed in. These events prevented a significant loss of life.

7. Positive outcome: Distal normal event: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During a storm, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges whose occupants were snowed in. These events prevented a significant loss of life.

8. Positive outcome: Distal abnormal event: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During one of the worst storms ever recorded, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche narrowly missed three full ski lodges whose occupants were snowed in. These events prevented a significant loss of life.

9. Negative outcome: Positive action: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A ski patroller knew he was risking his life, but he wanted to make sure the ski village at the bottom of the ski field was safe. The patroller climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

10. Negative outcome: Negative action: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A rebel wanted to avenge his brother's death by destroying the ski village at the bottom of the ski field. The rebel climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope

bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

11. Negative outcome: Distal normal event: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During a storm, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

12. Negative outcome: Distal abnormal event: Proximal normal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During one of the worst storms ever recorded, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, a clap of thunder resonated around the valley, setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

13. Negative outcome: Positive action: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A ski patroller knew he was risking his life, but he wanted to make sure the ski village at the bottom of the ski field was safe. The patroller climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

14. Negative outcome: Negative action: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. A rebel wanted to avenge his brother's death by destroying the ski village at the bottom of the ski field. The rebel climbed to the top of the mountain and deliberately blew up one of the largest cornices, making it roll down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

15. Negative outcome: Distal normal event: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During a storm, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

16. Negative outcome: Distal abnormal event: Proximal abnormal event.

This winter, record snowfalls accumulated creating many overhanging ledges and icy cornices on the steep ridges in the mountains. During one of the worst storms ever recorded, one of the largest cornices at the top of the mountain broke off and rolled down the slope bringing about an avalanche. Moments later, lightning struck a rocky outcrop triggering a massive rock-fall, the sound of which resonated around the valley setting off another avalanche. Just as the avalanches bore down on the village they converged, changing their paths. As a result, the avalanche swept into three full ski lodges whose occupants were snowed in. These events led to the loss of 73 lives.

Scenario: Flooding

1. Positive outcome: Distal positive action: Proximal normal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch

of levee. The engineer knew this would result in several local farmhouses being destroyed (one of which was his own), but he wanted to save the historical towns downstream. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns. Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

2. Positive outcome: Distal negative action: Proximal normal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch of levee. The engineer knew this would result in several local farmhouses being destroyed, but he choose that particular site as he wanted to get even with his ex-wife who lived in one houses. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns. Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

3. Positive outcome: Distal normal event: Proximal normal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns. Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

4. Positive outcome: Distal abnormal event: Proximal normal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, abnormally heavy rainfall saw once-in-100 year river levels being reached and a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns.

Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

5. Positive outcome: Distal positive action: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch of levee. The engineer knew this would result in several local farmhouses being destroyed (one of which was his own), but he wanted to save the historical towns downstream. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns. Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

6. Positive outcome: Distal negative action: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch of levee. The engineer knew this would result in several local farmhouses being destroyed, but he choose that particular site as he wanted to get even with his ex-wife who lived in one houses. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns. Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

7. Positive outcome: Distal normal event: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns. Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

8. Positive outcome: Distal abnormal event: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, abnormally heavy

rainfall saw once-in-100 year river levels being reached and a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events prevented the floodwaters from reaching the historical towns. Significant damage was reported in the surrounding farmland, but the 13,500 residents evacuated from the towns returned home to minimal damage.

9. Negative outcome: Distal positive action: Proximal normal event.

Rapidly rising floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch of levee. The engineer knew this would result in several local farmhouses being destroyed (one of which was his own), but he wanted to save the historical towns downstream. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events led to the floodwaters reaching the historical towns. Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

10. Negative outcome: Distal negative action: Proximal normal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch of levee. The engineer knew this would result in several local farmhouses being destroyed, but he choose that particular site as he wanted to get even with his ex-wife who lived in one houses. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events led to the floodwaters reaching the historical towns. Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

11. Negative outcome: Distal normal event: Proximal normal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events led to the floodwaters reaching the historical towns.

Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

12. Negative outcome: Distal abnormal event: Proximal normal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, abnormally heavy rainfall saw once-in-100 year river levels being reached and a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock was considered to be very unstable, a massive landslide occurred diverting the floodwaters. These events led to the floodwaters reaching the historical towns. Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

13. Negative outcome: Distal positive action: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch of levee. The engineer knew this would result in several local farmhouses being destroyed (one of which was his own), but he wanted to save the historical towns downstream. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events led to the floodwaters reaching the historical towns. Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

14. Negative outcome: Distal negative action: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. To alleviate pressure downstream, an engineer blew up a 200m stretch of levee. The engineer knew this would result in several local farmhouses being destroyed, but he choose that particular site as he wanted to get even with his ex-wife who lived in one houses. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events led to the floodwaters reaching the historical towns. Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

15. Negative outcome: Distal normal event: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events led to the floodwaters reaching the historical towns. Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

16. Negative outcome: Distal abnormal event: Proximal abnormal event.

Floodwaters along Frog River threatened to destroy hundreds of hectares of farmland and several historical towns. Although the area was prone to annual flooding, abnormally heavy rainfall saw once-in-100 year river levels being reached and a levee failed. The sudden surge in floodwaters destroyed several local farmhouses. Later that day, in an area where the riverbank's bedrock is considered to be very stable, a massive landslide occurred, diverting the floodwaters. These events led to the floodwaters reaching the historical towns. Significant damage was reported and 13,500 residents evacuated from the towns were unable to return home.

Study 5

Three new scenarios were added in Study 5. As in Studies 1-3, Study 5 used a mixed design with three between-subject variables: distal event (positive action, negative action, or abnormal physical event), outcome (positive or negative) and scenario (fire, train, factory, avalanche, car, casino, or soldier). The 2 (outcome) x 3 (distal event) design resulted in six versions of each scenario.

Scenario: Car Accident

1. Positive outcome: Distal positive action.

One unusually cold and stormy afternoon, John was in a hurry to get home before his wife. John was deliberately driving over the speed limit (53kmh in a 50kmh area). His goal was to get home in time to hide an anniversary present that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result, John hit a car that was coming from the other direction. These events resulted in John's car shunting the other car away from a child who was walking on the footpath at the time. The child would certainly have been seriously injured if John's car had not hit the other car out of the way.

2. Positive outcome: Distal negative action.

One unusually cold and stormy afternoon, John was in a hurry to get home before his wife. John was deliberately driving over the speed limit (53kmh in a 50kmh area) His goal was to get home in time to hide a vial of cocaine that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result, John hit a car that was coming from the other direction. These events resulted in John's car shunting the other car away from a child who was walking on the footpath at the time. The child would certainly have been seriously injured if John's car had not hit the other car out of the way.

3. Positive outcome: Distal abnormal event.

One unusually cold and stormy afternoon, John was in a hurry to get home before his wife. John had to take a detour home as his normal route was blocked by a tree felled by a lightning strike. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result, John hit a car that was coming from the

other direction. These events resulted in John's car shunting the other car away from a child who was walking on the footpath at the time. The child would certainly have been seriously injured if John's car had not hit the other car out of the way.

4. Negative outcome: Distal positive action.

One unusually cold and stormy afternoon, John was in a hurry to get home before his wife. John was deliberately driving over the speed limit (53kmh in a 50kmh area). His goal was to get home in time to hide an anniversary present that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result, John hit a car that was coming from the other direction. These events resulted in John's car shunting the other car into a child who was walking on the footpath at the time. The child suffered multiple serious injuries.

5. Negative outcome: Distal negative action.

One unusually cold and stormy afternoon, John was in a hurry to get home before his wife. John was deliberately driving over the speed limit (53kmh in a 50kmh area) His goal was to get home in time to hide a vial of cocaine that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result, John hit a car that was coming from the other direction. These events resulted in John's car shunting the other car into a child who was walking on the footpath at the time. The child suffered multiple serious injuries.

6. Negative outcome: Distal abnormal event.

One unusually cold and stormy afternoon, John was in a hurry to get home before his wife. John had to take a detour home as his normal route was blocked by a tree felled by a lightning strike. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result, John hit a car that was coming from the other direction. These events resulted in John's car shunting the other car into a child who was walking on the footpath at the time. The child suffered multiple serious injuries.

Scenario: Casino

1. Positive outcome: Distal positive action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose a number. John liked Mary so he put Mary's chip on number 18 because he knew the roulette wheel was rigged and that this number

would *win*. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to continue bouncing around the wheel. These events resulted in Mary winning \$1 million dollars.

2. Positive outcome: Distal negative action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose a number. John was jealous of Mary so he put Mary's chip on number 18 because he knew the roulette wheel was rigged and that this number would *lose*. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to continue bouncing around the wheel. These events resulted in Mary winning \$1 million dollars.

3. Positive outcome: Distal abnormal event.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary had just entered her number choice into the roulette ticket machine when a sudden power surge caused an override and the machine allocated her a new number. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to continue bouncing around the wheel. These events resulted in Mary winning \$1 million dollars.

4. Negative outcome: Distal positive action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose a number. John liked Mary so he put Mary's chip on number 18 because he knew the roulette wheel was rigged and that this number would *win*. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to continue bouncing around the wheel. These events resulted in Mary losing \$1 million dollars.

5. Negative outcome: Distal negative action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose a number. John was jealous of Mary so he put Mary's chip on number 18 because he knew the roulette wheel was rigged and that this number would *lose*. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to continue bouncing around the wheel. These events resulted in Mary losing \$1 million dollars.

6. Negative outcome: Distal abnormal event.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary had just entered her number choice into the roulette ticket machine when a sudden power surge caused an override and the machine allocated her a new number. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to continue bouncing around the wheel. These events resulted in Mary losing \$1 million dollars.

Scenario: Soldier

1. Positive outcome: Distal positive action.

In the middle of a war, a soldier whose identity is unknown is admitted into the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. The orderly knew to give the soldier a general antibiotic drug through an intra-venous drip (IV). The orderly checked he administered the correct dosage of the drug because he thought the soldier was an ally and he did not want him to die. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. Four hours later the soldier showed clear signs of improvement. Within 48 hours he had fully recovered with no adverse side effects.

2. Positive outcome: Distal negative action.

In the middle of a war, a soldier whose identity is unknown is admitted into the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. The orderly knew to give the soldier a general antibiotic drug through an intra-venous drip (IV). The orderly did not check he administered the correct dosage of the drug because he thought the soldier was an enemy and he wanted him to die. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. Four hours later the soldier showed clear signs of improvement. Within 48 hours he had fully recovered with no adverse side effects.

3. Positive outcome: Distal abnormal event.

In the middle of a war, a soldier whose identity is unknown is admitted into the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. A

general antibiotic drug was administered through the IV (intra-venous drip) to the soldier after an earthquake jolted the IV on/off switch to on. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. Four hours later the soldier showed clear signs of improvement. Within 48 hours he had fully recovered with no adverse side effects.

4. Negative outcome: Distal positive action.

In the middle of a war, a soldier whose identity is unknown is admitted into the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. The orderly knew to give the soldier a general antibiotic drug through an intra-venous drip (IV). The orderly checked he administered the correct dosage of the drug because he thought the soldier was an ally and he did not want him to die. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. Four hours later the soldier's condition rapidly deteriorated and he died shortly after.

5. Negative outcome: Distal negative action.

In the middle of a war, a soldier whose identity is unknown is admitted into the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. The orderly knew to give the soldier a general antibiotic drug through an intra-venous drip (IV). The orderly did not check he administered the correct dosage of the drug because he thought the soldier was an enemy and he wanted him to die. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. Four hours later the soldier's condition rapidly deteriorated and he died shortly after.

6. Negative outcome: Distal abnormal event.

In the middle of a war, a soldier whose identity is unknown is admitted into the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. A general antibiotic drug was administered through the IV (intra-venous drip) to the soldier after an earthquake jolted the IV on/off switch to on. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. Four hours later the soldier's condition rapidly deteriorated and he died shortly after.

Study 6

Several minor changes were made to the fire, train and factory scenarios across Studies 2-5, so a full set of scenarios used in Study 6 are provided below. Several changes were made to the car, soldier, and casino scenarios in Study 6. First, the action statement in the car scenario was revised to minimise the potential for normative violations influencing responses. For example, participants may have made a normative response to traffic violations e.g., *John was deliberately driving over the speed limit (53kmh in a 50kmh area)* and contextual assessment of John's driving skills given the poor weather conditions (the background condition described an '*unusually cold and stormy afternoon*'). Therefore, the explicit reference to exceeding the speed limit was deleted; the revised action statement read: *One unusually cold and stormy afternoon, John was in a hurry to get home before his wife.*

Second, it is possible that the distinct asymmetrical effects reported in the soldier scenario reflect action/in-action effects (Catellani, Alberici, & Milesi, 2004; Catellani & Milesi, 2001; Kahneman & Miller, 1986; Knobe, 2003), or perceived normative violations (Alicke et al., 2011) rather than motive valence. For example, *the orderly checks or does not check* the dosage rate was based on motive (positive or negative), whereas in all other scenarios the voluntary event is the same action. Second, participants may have perceived the orderly's decision to act based on the soldier's nationality as a violation of both social and professional norms (e.g., Geneva Conventions (1929, 1931, 1949) and the Hippocratic Oath). Therefore, the background statement and distal conditions were rewritten to eliminate these confounds.

Third, using the gaming structure in the casino scenario provided an opportunity to examine how people assess events in sequences that readily lend themselves to probabilistic predictions (e.g., Spellman, 1997). In the casino scenario, the probabilistic relationship between each event and the outcome can be considered as follows: before Mary plays roulette, in a game of double or nothing, the probability of winning is .5. After John places the bet, the rigged roulette wheel ensures the probability of winning is either 1 or 0. The proximal event (i.e., the earthquake jolt) causes the ball to continue bouncing but it does not alter 'the rigging' of the wheel and therefore the probability of the outcome does not change. However, it is possible that participants' 'intuitive probability assessment' may not have reflected the intended probability of a double or nothing win of .5, but assessed the probability of the outcome closer to that of a single number win that is, 1:37 or .02, because *John was asked to put Mary's chip on number 18*. To ensure

probabilistic equivalence, the distal event was revised to describe a 50:50 scenario (e.g., *Mary asked John to choose odd numbers or even numbers*).

In Study 6, the distal physical event in Studies 1-5 was replaced by an unstated motive, the action was the same in all motive conditions. Thus, Study 6 used a mixed design that included three between-subject variables: scenario (fire, train, factory, avalanche, car, casino, or soldier), outcome (positive or negative) and distal motive (positive action, negative action, or unstated motive action). The 2 (outcome) x 3 (distal motive) design resulted in six versions of each scenario.

Scenario: Forest Fire

1. Positive outcome: Distal positive action.

A forestry town was surrounded by wildfires and had been completely cut off from outside help. A young man deliberately set fire to an area of forest. His goal was to cause a burn-off to prevent the wildfires from reaching the town. That night a tornado struck, fanning the flames and uprooting several trees, diverting fire engines around blocked roads. As a result the wildfires moved away from the town and the lives of dozens of people trapped in their homes were saved.

2. Positive outcome: Distal negative action.

A forestry town was surrounded by wildfires and had been completely cut off from outside help. A young man deliberately set fire to an area of forest. His goal was to cause a second fire to ensure the wildfires reached the town. That night a tornado struck, fanning the flames and uprooting several trees, diverting fire engines around blocked roads. As a result the wildfires moved away from the town and the lives of dozens of people trapped in their homes were saved.

3. Positive outcome: Distal unstated motive action.

A forestry town was surrounded by wildfires and had been completely cut off from outside help. A young man deliberately set fire to an area of forest. That night a tornado struck, fanning the flames and uprooting several trees, diverting fire engines around blocked roads. As a result the wildfires moved away from the town and the lives of dozens of people trapped in their homes were saved.

4. Negative outcome: Distal positive action.

A forestry town was surrounded by wildfires and had been completely cut off from outside help. A young man deliberately set fire to an area of forest. His goal was to cause a burn-off to prevent the wildfires from reaching the town. That night a tornado struck, fanning the flames and

uprooting several trees, diverting fire engines around blocked roads. As a result the wildfires moved into the town and the lives of dozens of people trapped in their homes were lost.

5. Negative outcome: Distal negative action.

A forestry town was surrounded by wildfires and had been completely cut off from outside help. A young man deliberately set fire to an area of forest. His goal was to cause a second fire to ensure the wildfires reached the town. That night a tornado struck, fanning the flames and uprooting several trees, diverting fire engines around blocked roads. As a result the wildfires moved into the town and the lives of dozens of people trapped in their homes were lost.

6. Negative outcome: Distal unstated motive action.

A forestry town was surrounded by wildfires and had been completely cut off from outside help. A young man deliberately set fire to an area of forest. That night a tornado struck, fanning the flames and uprooting several trees, diverting fire engines around blocked roads. As a result the wildfires moved into the town and the lives of dozens of people trapped in their homes were lost.

Scenario: Train Accident

1. Positive outcome: Distal positive motive.

Just after 3:30pm, a full school bus approached a level crossing. Earlier that afternoon, a man deliberately changed the points at a junction on a railway track. His goal was to re-route the 3:32pm bullet train onto a second track to prevent a train collision. At 3:18pm, an earthquake caused a barrier arm to jam letting traffic through the railway crossing. As a result the bullet train flashed past the school bus that was on the level crossing at the time and the lives of dozens of schoolchildren were saved.

2. Positive outcome: Distal negative action.

Just after 3:30pm, a full school bus approached a level crossing. Earlier that afternoon, a man deliberately changed the points at a junction on a railway track. His goal was to re-route the 3:32pm bullet train onto a second track to cause a train collision. At 3:18pm, an earthquake caused a barrier arm to jam letting traffic through the railway crossing. As a result the bullet train flashed past the school bus that was on the level crossing at the time and the lives of dozens of schoolchildren were saved.

3. *Positive outcome: Distal unstated motive action.*

Just after 3:30pm, a full school bus approached a level crossing. Earlier that afternoon, a man deliberately changed the points at a junction on a railway track. At 3:18pm, an earthquake caused a barrier arm to jam letting traffic through the railway crossing. As a result the bullet train flashed past the school bus that was on the level crossing at the time and the lives of dozens of schoolchildren were saved.

4. *Negative outcome: Distal positive action.*

Just after 3:30pm, a full school bus approached a level crossing. Earlier that afternoon, a man deliberately changed the points at a junction on a railway track. His goal was to re-route the 3:32pm bullet train onto a second track to prevent a train collision. At 3:18pm, an earthquake caused a barrier arm to jam letting traffic through the railway crossing. As a result the bullet train collided with the school bus that was on the level crossing at the time, and the lives of dozens of school children were lost.

5. *Negative outcome: Distal negative action.*

Just after 3:30pm, a full school bus approached a level crossing. Earlier that afternoon, a man deliberately changed the points at a junction on a railway track. His goal was to re-route the 3:32pm bullet train onto a second track to cause a train collision. At 3:18pm, an earthquake caused a barrier arm to jam letting traffic through the railway crossing. As a result the bullet train collided with the school bus that was on the level crossing at the time, and the lives of dozens of school children were lost.

6. *Negative outcome: Distal unstated motive action.*

Just after 3:30pm, a full school bus approached a level crossing. Earlier that afternoon, a man deliberately changed the points at a junction on a railway track. At 3:18pm, an earthquake caused a barrier arm to jam letting traffic through the railway crossing. As a result the bullet train collided with the school bus that was on the level crossing at the time, and the lives of dozens of school children were lost.

Scenario: Factory

1. *Positive outcome: Distal positive action.*

At the beginning of the processing season, the Ministry of Health warned a large poultry factory that it would be closed if another unacceptably high Salmonella reading occurred. Later that week, an employee deliberately adjusted the thermostat on a conveyor belt. Her goal was to

ensure the chicken meat temperature fell within safety standards to decrease the risk of food poisoning. A short time later, sunlight refracted through an icy windowpane focused on the automatic control unit's temperature gauge changing all of the plant's thermostats. As a result the meat was processed at the correct temperature allowing the plant to start fully operating with 154 jobs being created.

2. Positive outcome: Distal negative action.

At the beginning of the processing season, the Ministry of Health warned a large poultry factory that it would be closed if another unacceptably high Salmonella reading occurred. Later that week, an employee deliberately adjusted the thermostat on a conveyor belt. Her goal was to ensure the chicken meat temperature rose above safety standards to increase the risk of food poisoning. A short time later, sunlight refracted through an icy windowpane focused on the automatic control unit's temperature gauge changing all of the plant's thermostats. As a result the meat was processed at the correct temperature allowing the plant to start fully operating with 154 jobs being created.

3. Positive outcome: Distal unstated motive action.

At the beginning of the processing season, the Ministry of Health warned a large poultry factory that it would be closed if another unacceptably high Salmonella reading occurred. Later that week, an employee deliberately adjusted the thermostat on a conveyor belt. A short time later, sunlight refracted through an icy windowpane focused on the automatic control unit's temperature gauge changing all of the plant's thermostats. As a result the meat was processed at the correct temperature allowing the plant to start fully operating with 154 jobs being created.

4. Negative outcome: Distal positive action.

At the beginning of the processing season, the Ministry of Health warned a large poultry factory that it would be closed if another unacceptably high Salmonella reading occurred. Later that week, an employee deliberately adjusted the thermostat on a conveyor belt. Her goal was to ensure the chicken meat temperature fell within safety standards to decrease the risk of food poisoning. A short time later, sunlight refracted through an icy windowpane focused on the automatic control unit's temperature gauge changing all of the plant's thermostats. As a result the meat was processed at the incorrect temperature forcing the plant to stop operating with 154 jobs being lost.

5. Negative outcome: Distal negative action.

At the beginning of the processing season, the Ministry of Health warned a large poultry factory that it would be closed if another unacceptably high Salmonella reading occurred. Later that week, an employee deliberately adjusted the thermostat on a conveyor belt. Her goal was to ensure the chicken meat temperature rose above safety standards to increase the risk of food poisoning. A short time later, sunlight refracted through an icy windowpane focused on the automatic control unit's temperature gauge changing all of the plant's thermostats. As a result the meat was processed at the incorrect temperature forcing the plant to stop operating with 154 jobs being lost.

6. Negative outcome: Distal unstated motive action.

At the beginning of the processing season, the Ministry of Health warned a large poultry factory that it would be closed if another unacceptably high Salmonella reading occurred. Later that week, an employee deliberately adjusted the thermostat on a conveyor belt. A short time later, sunlight refracted through an icy windowpane focused on the automatic control unit's temperature gauge changing all of the plant's thermostats. As a result the meat was processed at the incorrect temperature forcing the plant to stop operating with 154 jobs being lost.

Scenario: Avalanche

1. Positive outcome: Distal positive action.

This winter, record snowfalls trapped many people in their homes in the mountain villages and created extreme avalanche risk conditions. A man climbed to the top of the mountain and deliberately blew up one of the largest icy overhangs, making it roll down the slope bringing about an avalanche. His goal was to ensure the village at the bottom of the ski field was safe. Moments later, a massive rock-fall crashed down the valley, converging with the avalanche. As a result the mass of avalanche and rock-fall debris veered away from three full ski lodges and the lives of the 73 occupants who had been snowed in were saved.

2. Positive outcome: Distal negative action.

This winter, record snowfalls trapped many people in their homes in the mountain villages and created extreme avalanche risk conditions. A man climbed to the top of the mountain and deliberately blew up one of the largest icy overhangs, making it roll down the slope bringing about an avalanche. His goal was to ensure the village at the bottom of the ski field was destroyed. Moments later, a massive rock-fall crashed down the valley, converging with the

avalanche. As a result the mass of avalanche and rock-fall debris veered away from three full ski lodges and the lives of the 73 occupants who had been snowed in were saved.

3. Positive outcome: Distal unstated motive action.

This winter, record snowfalls trapped many people in their homes in the mountain villages and created extreme avalanche risk conditions. A man climbed to the top of the mountain and deliberately blew up one of the largest icy overhangs, making it roll down the slope bringing about an avalanche. Moments later, a massive rock-fall crashed down the valley, converging with the avalanche. As a result the mass of avalanche and rock-fall debris veered away from three full ski lodges and the lives of the 73 occupants who had been snowed in were saved.

4. Negative outcome: Distal positive action.

This winter, record snowfalls trapped many people in their homes in the mountain villages and created extreme avalanche risk conditions. A man climbed to the top of the mountain and deliberately blew up one of the largest icy overhangs, making it roll down the slope bringing about an avalanche. His goal was to ensure the village at the bottom of the ski field was safe. Moments later, a massive rock-fall crashed down the valley, converging with the avalanche. As a result the mass of avalanche and rock-fall debris veered into three full ski lodges and the lives of the 73 occupants who had been snowed in were lost.

5. Negative outcome: Distal negative action.

This winter, record snowfalls trapped many people in their homes in the mountain villages and created extreme avalanche risk conditions. A man climbed to the top of the mountain and deliberately blew up one of the largest icy overhangs, making it roll down the slope bringing about an avalanche. His goal was to ensure the village at the bottom of the ski field was destroyed. Moments later, a massive rock-fall crashed down the valley, converging with the avalanche. As a result the mass of avalanche and rock-fall debris veered into three full ski lodges and the lives of the 73 occupants who had been snowed in were lost.

6. Negative outcome: Distal unstated motive action.

This winter, record snowfalls trapped many people in their homes in the mountain villages and created extreme avalanche risk conditions. A man climbed to the top of the mountain and deliberately blew up one of the largest icy overhangs, making it roll down the slope bringing about an avalanche. Moments later, a massive rock-fall crashed down the valley, converging with

the avalanche. As a result the mass of avalanche and rock-fall debris veered into three full ski lodges and the lives of the 73 occupants who had been snowed in were lost.

Scenario: Casino

1. Positive outcome: Distal positive action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose odd numbers or even numbers. John deliberately put Mary's chip on 'evens'. His goal was to ensure Mary would win because he knew the roulette wheel was rigged and that an even number would win. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to bounce further around the wheel. As a result Mary won \$1 million dollars.

2. Positive outcome: Distal negative action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose odd numbers or even numbers. John deliberately put Mary's chip on 'evens'. His goal was to ensure Mary would lose because he knew the roulette wheel was rigged and that an odd number would win. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to bounce further around the wheel. As a result Mary won \$1 million dollars.

3. Positive outcome: Distal unstated motive action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose odd numbers or even numbers. John deliberately put Mary's chip on 'evens'. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to bounce further around the wheel. As a result Mary won \$1 million dollars.

4. Negative outcome: Distal positive action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose odd numbers or even numbers. John deliberately put Mary's chip on 'evens'. His goal was to ensure Mary would win because he knew the roulette wheel was rigged and that an even number would win. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to bounce further around the wheel. As a result Mary lost \$1 million dollars.

5. Negative outcome: Distal negative action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose odd numbers or even numbers. John deliberately put Mary's chip on 'evens'. His goal was to ensure Mary would lose because he knew the roulette wheel was rigged and that an odd number would win. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to bounce further around the wheel. As a result Mary lost \$1 million dollars.

6. Negative outcome: Distal unstated motive action.

Mary was at the casino and had \$1 million dollars. She was about to play a game of double or nothing at the roulette table. Mary asked John to choose odd numbers or even numbers. John deliberately put Mary's chip on 'evens'. The roulette wheel was spun. Just as the wheel stopped, an earthquake jolt caused the ball to bounce further around the wheel. As a result Mary lost \$1 million dollars.

Scenario: Car Accident

1. Positive outcome: Distal positive action.

One unusually cold and stormy afternoon, John was deliberately driving quickly to get home before his wife. His goal was to get home in time to hide an anniversary present that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result John hit a car that was coming from the other direction. John's car shunted the other car away from a child who was walking on the footpath at the time. The child would certainly have been seriously injured if John's car had not hit the other car out of the way.

2. Positive outcome: Distal negative action.

One unusually cold and stormy afternoon, John was deliberately driving quickly to get home before his wife. His goal was to get home in time to hide a vial of cocaine that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result John hit a car that was coming from the other direction. John's car shunted the other car away from a child who was walking on the footpath at the time. The child would certainly have been seriously injured if John's car had not hit the other car out of the way.

3. *Positive outcome: Distal unstated motive action.*

One unusually cold and stormy afternoon, John was deliberately driving quickly to get home before his wife. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result John hit a car that was coming from the other direction. John's car shunted the other car away from a child who was walking on the footpath at the time. The child would certainly have been seriously injured if John's car had not hit the other car out of the way.

4. *Negative outcome: Distal positive action.*

One unusually cold and stormy afternoon, John was deliberately driving quickly to get home before his wife. His goal was to get home in time to hide an anniversary present that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result John hit a car that was coming from the other direction. John's car shunted the other car into a child who was walking on the footpath at the time. The child suffered multiple serious injuries.

5. *Negative outcome: Distal negative action.*

One unusually cold and stormy afternoon, John was deliberately driving quickly to get home before his wife. His goal was to get home in time to hide a vial of cocaine that he had left out in the open before his wife could see it. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result John hit a car that was coming from the other direction. John's car shunted the other car into a child who was walking on the footpath at the time. The child suffered multiple serious injuries.

6. *Negative outcome: Distal unstated motive action.*

One unusually cold and stormy afternoon, John was deliberately driving quickly to get home before his wife. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because hail had made the road icy. As a result John hit a car that was coming from the other direction. John's car shunted the other car into a child who was walking on the footpath at the time. The child suffered multiple serious injuries.

Scenario: Soldier

1. *Positive outcome: Distal positive action.*

In the middle of a war, a soldier whose identity is unknown is taken to the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The

hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. While waiting for a hospital doctor, a nurse following approved procedures inserted an IV (intra-venous drip). A civilian, who works in a research laboratory, deliberately administered a drug through the IV. Her goal was to try and save the soldier's life because she thought he was an ally. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. As a result four hours later the soldier's condition rapidly improved and he fully recovered with no adverse side effects.

2. Positive outcome: Distal negative action.

In the middle of a war, a soldier whose identity is unknown is taken to the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. While waiting for a hospital doctor, a nurse following approved procedures inserted an IV (intra-venous drip). A civilian, who works in a research laboratory, deliberately administered a drug through the IV. Her goal was to try and kill the soldier because she thought he was an enemy. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. As a result four hours later the soldier's condition rapidly improved and he fully recovered with no adverse side effects.

3. Positive outcome: Distal unstated motive action.

In the middle of a war, a soldier whose identity is unknown is taken to the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. While waiting for a hospital doctor, a nurse following approved procedures inserted an IV (intra-venous drip). A civilian, who works in a research laboratory, deliberately administered a drug through the IV. As a result four hours later the soldier's condition rapidly improved and he fully recovered with no adverse side effects.

4. Negative outcome: Distal positive action.

In the middle of a war, a soldier whose identity is unknown is taken to the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. While waiting for a hospital doctor, a nurse following approved procedures inserted an IV (intra-venous drip). A civilian, who works in a research laboratory, deliberately administered

a drug through the IV. Her goal was to try and save the soldier's life because she thought he was an ally. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. As a result four hours later the soldier's condition rapidly deteriorated and he died shortly after.

5. Negative outcome: Distal negative action.

In the middle of a war, a soldier whose identity is unknown is taken to the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. While waiting for a hospital doctor, a nurse following approved procedures inserted an IV (intra-venous drip). A civilian, who works in a research laboratory, deliberately administered a drug through the IV. Her goal was to try and kill the soldier because she thought he was an enemy. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. As a result four hours later the soldier's condition rapidly deteriorated and he died shortly after.

6. Negative outcome: Distal unstated motive action.

In the middle of a war, a soldier whose identity is unknown is taken to the battlefield hospital. He is critically ill and suffering from a lung disease that invariably leads to death. The hospital doctors were busy with other critical casualties and were unable to attend the unknown soldier. While waiting for a hospital doctor, a nurse following approved procedures inserted an IV (intra-venous drip). A civilian, who works in a research laboratory, deliberately administered a drug through the IV. A short time later, an electrical storm caused a power surge resulting in the IV's control unit changing the dosage rate. As a result four hours later the soldier's condition rapidly deteriorated and he died shortly after.

Appendix B: Study 1 – Descriptive Statistics

Table B1

Means, and SDs for Cause, Blame, and Voluntariness Ratings for Distal and Proximal Events by Condition (Study 1)

Variable	Cause		Blame		Voluntariness	
	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome
Distal Event						
Positive Action	60.00 (29.92)	44.29 (31.55)	49.55 (36.05)	38.10 (32.81)	69.55 (37.98)	72.38 (37.00)
Negative Action	49.55 (39.09)	78.70 (23.99)	56.36 (41.24)	79.57 (22.46)	95.00 (9.67)	80.00 (29.33)
Physical Event	59.00 (29.18)	60.83 (32.73)	40.00 (32.97)	45.00 (35.87)	53.16 (37.31)	12.86 (25.23)
Proximal Event						
Positive Action	49.09 (32.06)	73.33 (24.97)	44.55 (36.61)	69.52 (34.13)	23.18 (34.14)	19.05 (32.70)
Negative Action	74.09 (25.38)	44.35 (32.59)	57.73 (28.61)	32.61 (32.08)	21.82 (38.25)	19.05 (30.81)
Physical Event	59.00 (30.07)	57.50 (29.08)	57.00 (27.93)	61.25 (32.88)	52.11 (40.22)	22.86 (33.19)

Appendix C: Study 2 – Descriptive Statistics

Table C1

Means, and SDs for Cause, Praise/Blame, and Responsibility Ratings for Distal and Proximal Events by Condition (Study 2)

Variable	Cause		Praise / Blame		Responsibility	
	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Distal Event						
Positive Action	56.81 (30.25)	36.43 (32.75)	44.09 (27.05)	36.91 (25.90)	59.55 (31.47)	44.05 (38.00)
Negative Action	59.76 (35.08)	76.63 (25.34)	82.33 (22.66)	84.09 (20.94)	61.86 (35.54)	82.50 (20.47)
Physical Event	73.90 (26.82)	64.00 (24.25)	69.51 (24.09)	65.11 (24.09)	54.14 (35.71)	52.22 (28.03)
Proximal Event						
Positive Action	61.60 (28.28)	66.43 (32.69)	48.63 (32.61)	79.52 (20.24)	55.91 (36.81)	67.85 (35.44)
Negative Action	54.42 (33.83)	62.73 (29.36)	36.74 (28.01)	65.22 (22.46)	60.23 (33.56)	58.68 (30.29)
Physical Event	41.22 (36.27)	59.11 (30.44)	45.37 (32.10)	68.22 (23.96)	59.51 (33.61)	62.90 (31.99)

Table C2

Means, and SEs for Within-Subjects Variable Judgment (Cause, Praise/Blame and Responsibility) for Distal and Proximal events with Positive and Negative outcomes (Study 2)

Variable	Positive Outcome		Negative Outcome	
	Distal Event	Proximal Event	Distal Event	Proximal Event
Positive Action	53.49	55.38	39.13	71.27
	(3.38)	(3.57)	(3.46)	(3.65)
Negative Action	67.98	50.46	80.08	62.27
	(3.42)	(3.61)	(3.38)	(3.57)
Physical Event	65.85	48.70	60.44	63.19
	(3.50)	(3.70)	(3.34)	(3.53)

Table C3

Means, and SDs for the difference between the proximal and distal events in the causal chain for Cause, Praise/Blame, and Responsibility by condition (Study 2)

Variable	Cause		Praise/Blame		Responsibility	
	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Positive Action	4.77	30.00	4.54	42.62	-3.64	23.81
	(51.78)	(55.08)	(47.76)	(40.00)	(59.42)	(65.74)
Negative Action	-5.35	-10.91	-45.58	-18.86	-1.63	-23.64
	(61.85)	(44.86)	(41.65)	(34.45)	(58.59)	(41.77)
Physical Event	-32.68	-4.89	-24.15	3.11	5.37	10.00
	(50.74)	(45.81)	(48.57)	(36.97)	(52.83)	(43.12)

Appendix D: Study 2 –Relations between Judgments of Cause, Blame, and Responsibility

The two 3-way interactions with judgment reported in Chapter Three²⁰ show that judgments of cause, praise/blame, and responsibility differ, but not exactly as predicted. A MANOVA comparing ratings across positive and negative outcomes clarified the 3-way interaction between outcome, temporal position and judgment. As predicted, there was a multivariate main effect for outcome, $F(6, 252) = 12.53, p < .001, \eta^2 = .23$, where judgments were amplified for negative rather than positive outcomes. However, follow-up univariate ANOVAs showed contrary to predictions, negative outcomes did not amplify ratings of cause, responsibility or praise/blame for distal events, but did amplify ratings for proximal physical for judgments of cause, $F(1, 257) = 6.32, p < .02, \eta^2 = .02$, and praise/blame, $F(1, 257) = 64.41, p < .001, \eta^2 = .20$, but not responsibility.

Two separate ANOVAs for distal and proximal events clarified the 3-way interaction between distal event, temporal position and judgment. The ANOVAs were performed with between subjects variable: 3 (distal event), and the within-subjects variable: 3 (judgments) for ratings of cause, praise/blame, and responsibility. See Appendix C1 for descriptive statistics. There was a 2-way interaction between distal event and judgments for ratings of distal events, $F(4, 514) = 13.74, p < .001, \eta^2 = .10$, and proximal events, $F(4, 514) = 3.07, p < .02, \eta^2 = .02$.

Follow-up ANOVAs showed that participants differentiated between the constructs of cause, praise/blame and responsibility, but not exactly as predicted for positive actions, $F(2, 85) = 4.76, p < .01, \eta^2 = .10$, negative actions, $F(2, 85) = 16.34, p < .001, \eta^2 = .28$, and distal physical events, $F(2, 84) = 11.24, p < .001, \eta^2 = .21$). Within-subject contrasts showed, as predicted, the valence of the action had a stronger effect on judgments of praise/blame than responsibility; praise/blame ratings were *higher* than responsibility ratings for negative actions, but *lower* than responsibility ratings for positive actions. Furthermore, the type of event (voluntary or physical) rather than the valence of the action had the stronger effect on judgments of cause and responsibility; ratings of cause and responsibility did not differ for positive or negative actions, but cause ratings were higher than responsibility ratings for distal physical events. Contrary to predictions, there was no difference between cause or praise/blame ratings for positive actions or distal physical events, whereas cause ratings were lower than blame ratings for negative actions.

²⁰ The two 3-way interactions were between outcome, temporal position, and judgment, $F(2, 252) = 8.00, p < .001, \eta^2 = .06$, and distal event, temporal position and judgment, $F(4, 506) = 10.99, p < .001, \eta^2 = .08$.

In contrast to ratings for distal events, participants did not distinguish between the constructs for ratings of proximal physical events when they followed positive or negative actions, but did when they followed physical events, $F(2, 84) = 4.27, p < .02, \eta^2 = .09$. Within-subject contrasts showed in chains of two physical events, there was no difference between praise/blame and cause or responsibility ratings, but cause ratings were lower than responsibility ratings.

Appendix E: Study 3 – Descriptive Statistics

Table E1

Means, and SDs for Cause, Praise/Blame, Responsibility and Punitiveness Ratings for Distal and Proximal Events by Condition (Study 3)

Variable	Cause		Praise / Blame		Responsibility		Punitiveness	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Distal Event								
Positive Action	62.78	43.89	36.94	52.78	72.78	47.22	40.00	54.44
	(27.32)	(26.11)	(23.40)	(22.25)	(25.03)	(32.83)	(18.82)	(15.20)
Negative Action	55.28	72.29	80.56	86.29	665.30	79.14	82.78	80.57
	(36.45)	(22.11)	(27.04)	(20.30)	(34.24)	(26.28)	(26.25)	(24.61)
Physical Event	64.17	61.94	57.22	66.94	60.56	57.22	54.17	57.22
	(26.12)	(23.03)	(21.46)	(18.02)	(32.68)	(27.21)	(18.88)	(14.06)
Proximal Event								
Positive Action	58.61	77.78	54.72	73.89	54.17	71.11	57.78	56.67
	(29.29)	(16.41)	(21.97)	(20.46)	(28.82)	(25.39)	(13.55)	(20.00)
Negative Action	65.00	61.14	42.78	66.86	60.28	52.29	48.06	60.28
	(32.21)	(22.72)	(20.50)	(19.52)	(33.42)	(26.13)	(17.54)	(20.79)
Physical Event	51.94	67.78	50.28	75.56	61.11	58.61	55.56	67.50
	(28.77)	(21.26)	(20.35)	(20.49)	(30.69)	(28.30)	(17.96)	(17.95)

Table E2

Mean and SDs for the Difference Between the Proximal and Distal Events in the Causal Chain for Cause, Praise/Blame, Responsibility, Punitiveness and Probability ratings by condition (Study 3)

Variable	Cause		Responsibility		Praise / Blame		Punitiveness		Probability	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Positive Action	-4.17	33.89	-18.61	23.89	17.78	21.11	17.78	2.22	11.75	23.58
	(43.38)	(35.56)	(38.56)	(49.64)	(36.57)	(31.42)	(25.42)	(25.65)	(77.29)	(61.11)
Negative Action	9.72	-11.15	-6.11	-26.85	-37.78	-19.43	-34.72	-20.29	27.81	-9.34
	(61.39)	(32.79)	(57.23)	(34.79)	(37.95)	(23.88)	(32.56)	(22.81)	(85.18)	(41.33)
Physical Event	-12.22	5.83	0.56	1.39	-6.94	8.61	1.39	10.28	34.58	35.56
	(41.27)	(33.24)	(41.26)	(33.05)	(27.45)	(21.27)	(27.37)	(20.91)	(84.50)	(49.47)

Appendix F: Study 4 - Pilot Studies

Pilot Study 1: The influence of two different types of motivation on attributions

Pilot Study 1 investigated whether participants would differentiate between two types of motivation: instrumental motivation (e.g., personal reward or personal cost), and situational motivation (e.g., self-defense or employment conditions). First, based on research (Reeder et al., 2002), I predicted that ratings would be higher for actions motivated for instrumental (e.g., knowingly risk life to save others) rather than situational reasons (e.g., lighting a burn-off).

Method

Design and participants.

There were three between-subject variables: 2 distal action (positive or negative), 2 outcome (positive or negative), and 2 type of motivation (instrumental or situational). Participants were voluntarily recruited on campus from Victoria University of Wellington ($N = 140$). Each participant was given a confectionary in appreciation of their participation.

Scenarios and procedure.

Three new scenarios were developed. The car scenario was adapted from Alicke (1992) and the avalanche scenario was adapted from Hilton et al., (2009). The flood scenario was new. Replicating Studies 1 – 3, each scenario depicted a sequence of two events leading to either a positive or negative outcome. The motive statement in the avalanche, car and flood scenarios aligned with instrumental motivation reasons, whereas the motive statement in the fire, factory and train scenarios described situational reasons. The proximal event was the same physical event for all conditions in each scenario.

Participants were presented with one of six scenarios (fire, train, factory, car, avalanche, flooding). Each participant read a single scenario followed by the measure of goodness/badness. The measure of goodness/badness replicated Alicke et al.'s (2011, Expt. 2) evaluative measure and read: 'Ignoring the outcome, how good or bad do you think the [Conservation Department officer's] action was?' Ratings were made on a Likert scale, where 0 = "Very bad" and 10 = "Very good".

Results and Discussion

An ANOVA was conducted with three between-subjects variables: 2 (outcome: positive, negative) x 2 (distal event: positive action, negative action) x 2 (motive-type: instrumental, situational) for evaluative judgments (good or bad). See Table F1 for descriptive statistics.

Table F1

Mean and SDs for Goodness or Badness ratings for distal events by condition (Pilot 1)

Variable	Positive Outcome		Negative Outcome	
	Situational	Instrumental	Situational	Instrumental
Distal Event				
Positive Action	6.33 (2.95)	6.11 (2.52)	6.88 (2.75)	5.33 (2.30)
Negative Action	1.74 (2.49)	2.59 (2.62)	0.59 (0.87)	0.88 (1.58)

The analyses revealed a main effect for outcome, $F(1, 132) = 3.75, p < .05, \eta^2 = .03$, where actions were rated more negatively in chains with negative outcomes ($M = 3.39, SE = 0.28$) than positive outcomes ($M = 4.19, SE = 0.27$). Verifying the experimental manipulation, there was also a main effect for distal event-type, $F(1, 132) = 139.69, p < .001, \eta^2 = .51$, where negatively motivated actions were very bad ($M = 1.43, SE = 0.28$), and positively motivated actions were good ($M = 6.16, SE = 0.28$). There were no interactions or main effect for type of motive.

Contrary to predictions, these results showed that exaggerating the instrumental properties of the action (i.e., the personal cost or risk to the actor) did not amplify participants' ratings; actions performed with negative intent were judged very bad. Interestingly however, positive actions describing heroic deeds or actions performed with high personal costs were seen no more positively than those performed as part of their role.

Pilot Study 2: The influence of abnormality, and two types of physical events on attributions.

Pilot Study 2 investigated firstly, whether the abnormality of the physical event influenced judgments and secondly, whether participants differentiated between mechanical and natural events. Based on legal theory and research (Hitchcock & Knobe, 2009; Lombrozo, 2010), it was predicted that abnormal events would be seen as better explanations of the outcome than normal events, and second, that mechanical events would be seen as better explanations of the outcome than natural events.

Method

Design and participants.

There were four between-subject variables: 2 distal physical event (normal or abnormal), 2 proximal event (natural or mechanical), 2 outcome (positive or negative) and 6 scenario (fire, train, factory, flood, avalanche and car) and one within-subject variable: 2 temporal position (distal event or proximal event). Participants were voluntarily recruited on campus from Victoria University of Wellington ($N = 140$). Each participant was given a confectionary in appreciation of their participation.

Measures and procedure.

Two revisions were made to the scenarios used in Pilot Study 1. First, the distal physical event (which was the same physical event for all conditions, e.g., a lightning strike set fire to an area of forest) was adapted to include two versions; a normal physical event and an abnormal physical event. The physical event statement was prefaced with a phrase that described the normal environment. The abnormal physical event was designed to meet legal criteria for abnormal physical events (Moore, 2009). Thus, the abnormality of the distal physical event was manipulated with the word '*frequently*' or '*rarely*'.

Second, the proximal physical event was adapted to describe two non-voluntary conditions that were equivalent in terms of their effects on the outcome; a natural physical event or a mechanical event. The physical event was revised to meet abnormality criteria, for example 'the strong wind' was replaced with 'a tornado'.

Each participant was given a single scenario followed by a series of questions that elicited judgments of the distal and proximal events' goodness of explanation and abnormality on a rating scale. The measure of goodness of explanation replicated that used by McClure et al., (2007). The measure of goodness of explanation read: How good is the statement [*distal or proximal event*] as an explanation of the [*outcome*]? Ratings of each event were made on a Likert scale where 0 = Not at all good, and 10 = Extremely good explanation.

The measure of abnormality replicated that used by Ryder, (unpublished 1996, supervisor J. McClure). The measure of abnormality read 'Please rate the extent to which you think each of the following events is unusual'. The stem was followed by the distal physical event e.g., *The lightning strike setting fire to an area of forest*, and the second physical event e.g., *The tornado fanning the flames*. Ratings of each event were made on a rating scale, where 0 = very unusual and 10 = Not at all unusual.

Results and Discussion

A MANOVA was conducted with three between-subjects variables: 2 (distal physical event-type: normal, abnormal) x 2 (outcome: positive, negative) x 6 (scenario: fire, train, factory, flood, avalanche, car) for judgments of abnormality and goodness of explanation. Descriptive statistics are shown in Table F2.

Table F2

Mean and SDs for Abnormality and Goodness of Explanation Ratings for Distal and Proximal Events by Condition (Pilot 2)

Variable	Abnormality		Good Explanation	
	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome
Distal Event				
Normal	5.17 (2.84)	5.17 (2.84)	3.22 (2.54)	4.19 (2.55)
Abnormal	5.36 (3.45)	5.36 (3.45)	2.69 (2.48)	5.00 (3.03)
Proximal Event				
Natural Event	4.47 (2.74)	4.47 (2.74)	4.69 (2.67)	4.53 (2.54)
Mechanical Event	4.92 (3.44)	4.92 (3.44)	3.56 (2.92)	5.15 (2.83)

There was a multivariate main effect for outcome, $F(1, 116) = 7.32, p < .001, \eta^2 = .11$. Follow-up univariate analyses showed ratings were higher for negative rather than positive outcomes for judgments of goodness of explanation, $F(1, 117) = 13.82, p < .001, \eta^2 = .11$, but not abnormality, $F(1, 116) = 0.94, ns$. There were no main effects or interactions for scenario or distal physical event-type. Thus, contrary to predictions participants did not see abnormal physical events as more unusual or better explanations for the outcome than normal physical events.

Natural physical events vs. mechanical events.

A MANOVA was conducted with three between-subjects variables: 2 x (proximal event-type: natural, mechanical), x 2 (outcome: positive, negative) x 6 (scenario: fire, train, factory, flood, avalanche, car) for judgments of goodness of explanation and abnormality.

There was a multivariate main effect for scenario, $F(10, 234) = 2.00, p < .04, \eta^2 = .08$. so separate analyses were performed for positive and negative outcomes. Follow-up univariate analyses showed the scenario effect was significant for judgments of abnormality, $F(5, 117) = 3.52, p < .01, \eta^2 = .13$, but not goodness of explanation, $F(5, 117) = 0.74, ns$. Post hoc boneferroni ($p < .05$) tests showed participants saw the mechanical event in the car scenario ($M = 6.58, SE = 0.62$) as less unusual than the mechanical event in the fire scenario ($M = 3.13, SE = 0.62$) or the natural event in the factory scenario ($M = 3.91, SE = 0.62$). There were no differences between any other scenarios. There were no main effects or interactions for outcome or proximal physical event-type. Thus, contrary to Lombrozo (2010) and Hitchcock and Knobe's (2009) findings, mechanical events were not rated better explanations of the outcome than natural physical events.

Appendix G: Study 4 – Descriptive Statistics

Table G1

Means and SDs for Abnormality, Cause, Responsibility, Praise/Blame and Punitiveness Ratings for Distal and Proximal Events by Condition (Study 4)

Variable	Abnormality		Cause		Responsibility		Praise / Blame		Punitiveness	
	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome
Distal Event										
Positive Action	43.88 (26.50)	40.43 (28.61)	54.22 (28.90)	43.74 (26.51)	58.71 (27.14)	47.39 (29.08)	44.66 (22.24)	51.48 (21.16)	44.83 (18.63)	53.74 (15.24)
Negative Action	67.20 (26.86)	66.81 (25.40)	49.24 (32.48)	69.45 (26.54)	54.58 (34.29)	75.75 (25.91)	77.63 (21.11)	86.64 (17.61)	78.98 (22.91)	84.87 (18.18)
Normal Physical	42.55 (26.52)	43.45 (27.68)	55.73 (26.04)	58.50 (25.26)	57.18 (26.86)	64.96 (20.92)	60.00 (16.59)	68.14 (15.56)	54.00 (15.98)	59.47 (15.23)
Abnormal Physical	56.06 (27.25)	57.26 (25.15)	50.28 (29.01)	58.32 (23.37)	48.62 (30.63)	54.69 (25.29)	55.96 (14.15)	62.74 (15.77)	52.84 (12.03)	53.98 (11.54)
Proximal Physical Event										
Normal Physical	54.45 (30.50)	51.03 (28.64)	64.15 (25.87)	66.07 (23.65)	59.43 (28.19)	62.63 (26.93)	48.08 (17.77)	63.97 (18.09)	49.21 (13.12)	55.27 (12.98)
Abnormal Physical	59.78 (28.48)	57.65 (29.16)	66.07 (23.77)	66.57 (21.45)	63.75 (26.25)	63.91 (23.94)	45.22 (16.99)	63.35 (17.17)	47.46 (12.61)	52.91 (11.63)

Table G2.

Means, and SDs for Abnormality, Cause, Responsibility, Praise/Blame and Punitiveness Ratings for Type of Chain by Condition (Study 4)

Variable	Abnormality		Cause		Responsibility		Praise/blame		Punitiveness	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Positive action:	14.00	9.81	9.83	20.55	3.17	15.64	3.17	7.09	6.33	0.55
Normal event	(43.03)	(50.09)	(41.56)	(39.08)	(37.08)	(44.63)	(20.71)	(30.65)	(20.33)	(22.15)
Positive action:	12.14	17.17	10.18	35.50	1.25	22.17	-1.25	17.67	-1.61	2.33
Abnormal event	(38.69)	(44.91)	(36.16)	(36.24)	(38.05)	(47.12)	(26.67)	(27.52)	(23.10)	(15.77)
Negative action:	-14.75	-22.73	23.77	-8.00	13.11	-22.55	-32.79	-24.91	-32.62	-32.18
Normal event	(35.22)	(42.18)	(42.82)	(39.88)	(55.15)	(43.43)	(22.15)	(27.48)	(22.87)	(15.12)
Negative action:	-13.68	-13.97	17.89	-9.96	1.05	-11.90	-30.70	-23.79	-29.30	-30.00
Abnormal event	(45.73)	(40.65)	(40.26)	(31.89)	(40.91)	(34.31)	(28.69)	(24.34)	(35.10)	(27.45)
Normal physical:	17.04	11.61	12.59	10.89	10.19	5.36	-10.37	1.96	-4.81	0.00
Normal event	(33.23)	(25.14)	(37.02)	(31.00)	(33.67)	(25.22)	(23.15)	(10.86)	(19.50)	(9.53)
Normal physical:	14.64	23.68	5.54	7.89	4.82	1.93	-9.46	-5.96	-2.86	-8.07
Abnormal event	(34.69)	(32.32)	(26.28)	(21.86)	(30.45)	(23.03)	(25.11)	(15.10)	(12.17)	(12.74)
Abnormal physical:	4.26	-6.38	7.96	5.34	10.74	4.48	-10.74	0.52	-7.41	-2.41
Normal event	(26.88)	(27.19)	(38.43)	(21.46)	(39.57)	(26.17)	(22.39)	(15.83)	(20.02)	(8.02)
Abnormal physical:	4.36	-0.55	13.27	7.09	9.64	4.91	-10.91	-1.82	-2.36	-1.82
Abnormal event	(37.65)	(31.76)	(36.26)	(30.04)	(33.39)	(23.72)	(18.88)	(15.99)	(12.17)	(11.88)

Appendix H: Study 4 - Tests of Legal Theories of Causation.

Targeted analyses tested Hart and Honoré's (1985) claims that proximal abnormal physical events would be rated higher than distal intentional actions as discussed in Chapter Four, Study 4. A mixed design ANOVA performed with two between-subjects variables: 2 (outcome) x 8 (chain-type) and one within-subjects variable: 2 (temporal position) for judgments of cause a 3-way interaction between outcome, chain-type and temporal position, $F(7, 891) = 8.25, p = .001, \eta^2 = .06$, so separate analyses were performed for positive and negative outcomes.

With negative outcomes, there was 2-way interaction between chain-type and temporal position, $F(7, 446) = 12.06, p = .001, \eta^2 = .16$. In chains with voluntary actions, the valence of the agent's motive rather than abnormality determined when proximal physical events were judged more causal than actions. Positive actions were rated *lower* than abnormal, $t(59) = 7.59, p < .001$, and normal, $t(54) = 3.90, p < .001$, proximal physical events, whereas negative actions were rated *higher* than abnormal physical events, $t(57) = 2.31, p < .03$, but did not differ from normal proximal events. In the baseline condition with two physical events, proximal events are rated *higher* than distal events for chains with two normal events, $t(55) = 2.63, p < .01$, and when abnormal events followed normal events, $t(56) = 2.73, p < .01$. In contrast, in chains with two abnormal events, $t(54) = 1.75, p < .09$, or when normal events followed abnormal events, $t(57) = 1.90, p < .07$, causal ratings did not differ. With positive outcomes, there was a main effect for temporal position, where proximal physical events were rated higher than distal events, $F(1, 445) = 50.59, p = .001, \eta^2 = .10$. There was no main effect for the type of chain or interaction.

These results do not support Hart and Honoré's (1985) claim that when chains comprise voluntary actions and physical events, people prefer the most recent voluntary action, unless that action is followed by an independent abnormal physical event. These results also counter the claim that when chains comprise two equivalent events (as in the baseline condition), people prefer the most recent event. Instead, these results show that with positive outcomes, regardless of voluntariness, the abnormality of the event, or the agent's motive, the most recent (proximal) event is seen as more causal than the distal event. With negative outcomes, however, the valence of the agent's motive rather than abnormality or proximity determines when physical events are rated more causal than actions.

Appendix I: Study 5 – Descriptive Statistics

Table I1

Means and SDs for Cause and Praise/Blame Ratings for Distal and Proximal Events by Condition (Study 5)

Variable	Cause		Praise / Blame	
	Positive Outcome	Negative Outcome	Positive Outcome	Negative Outcome
Distal Event				
Positive Action	55.33 (24.11)	42.92 (31.95)	43.60 (19.70)	80.35 (20.11)
Negative Action	46.51 (30.63)	58.28 (29.43)	65.48 (21.72)	65.00 (16.09)
Physical Event	56.62 (27.14)	56.47 (26.27)	48.61 (17.44)	67.02 (24.30)
Proximal Event				
Positive Action	58.72 (23.67)	70.33 (21.65)	51.40 (16.14)	71.67 (21.99)
Negative Action	68.17 (23.54)	57.16 (27.55)	42.06 (24.80)	62.76 (18.02)
Physical Event	55.31 (26.99)	61.03 (25.71)	47.05 (21.82)	66.90 (18.42)

Table I2

Means and SDs for Probabilistic Contribution Ratings for Background, Distal, and Proximal Events by Condition (Study 5)

Variable	Background	Distal event	Proximal event
Positive Action			
Positive Outcome	30.04 (21.65)	19.53 (29.42)	15.93 (30.85)
Negative Outcome	41.00 (28.13)	-0.85 (38.39)	36.52 (41.12)
Negative Action			
Positive Outcome	45.13 (32.45)	-13.38 (40.85)	33.62 (49.77)
Negative Outcome	31.38 (26.87)	20.52 (29.32)	29.40 (30.58)
Physical Event			
Positive Outcome	52.61 (32.42)	-9.54 (35.49)	19.66 (34.85)
Negative Outcome	32.64 (30.26)	14.05 (27.86)	22.10 (35.75)

Appendix J: Study 6 – Descriptive Statistics

Table J1

Means, and SDs for Cause, Praise/Blame and Responsibility Ratings for Distal and Proximal Events by Condition (Study 6)

Variable	Cause		Praise / Blame		Responsibility	
	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Distal Action						
Positive Motive	60.71 (19.79)	45.50 (26.34)	44.29 (23.18)	54.43 (21.29)	64.36 (25.26)	47.00 (29.53)
Negative Motive	54.12 (30.01)	66.50 (24.84)	72.54 (22.83)	84.79 (17.91)	56.30 (30.75)	74.29 (29.16)
Unstated Motive	52.57 (26.78)	52.35 (26.48)	51.99 (26.09)	74.49 (23.65)	58.09 (28.35)	62.35 (30.35)
Proximal Event						
Positive Motive	59.60 (22.42)	66.50 (27.95)	44.43 (20.88)	71.57 (22.52)	54.83 (26.67)	62.79 (29.58)
Negative Motive	71.52 (28.02)	54.79 (28.02)	39.20 (23.27)	61.29 (19.20)	66.81 (26.84)	49.14 (30.55)
Unstated Motive	65.00 (25.55)	66.62 (12.19)	41.03 (20.47)	67.06 (22.61)	63.38 (26.51)	60.81 (26.30)

Table J2

Means, and SDs for Belief, Desire, Intentionality, and Unitisation Ratings for Distal Actions by Condition (Study 6)

Variable	Intentionality		Belief		Desire		Unitisation of Chain	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Positive Motive	40.50 (37.89)	10.57 (15.92)	50.57 (36.63)	17.21 (25.65)	72.57 (34.40)	7.50 (20.69)	38.43 (35.68)	54.93 (34.89)
Negative Motive	18.91 (27.56)	52.43 (36.10)	18.12 (27.76)	55.07 (33.43)	12.61 (24.49)	61.07 (38.08)	41.67 (37.76)	56.71 (33.09)
Unstated Motive	34.19 (33.51)	17.94 (20.83)	36.19 (34.43)	17.57 (22.52)	59.78 (32.75)	18.01 (22.50)	61.10 (36.08)	45.51 (35.61)

Table I3

Means, and SDs for Objective, Subjective and Reasonable Foresight Ratings for Distal Actions by Condition (Study 6)

Variable	Objective Foresight		Subjective Foresight		Reasonable Foresight	
	Positive	Negative	Positive	Negative	Positive	Negative
	Outcome	Outcome	Outcome	Outcome	Outcome	Outcome
Positive action	21.54 (34.78)	1.79 (30.13)	29.07 (25.47)	14.77 (14.86)	55.04 (25.71)	36.71 (24.06)
Negative action	-13.55 (38.12)	24.71 (36.58)	11.52 (15.75)	45.07 (30.43)	33.04 (27.24)	55.14 (26.92)
Unstated action	-2.87 (35.31)	8.90 (34.42)	25.99 (27.60)	20.77 (18.75)	40.22 (21.38)	43.90 (24.90)

Table 1.1. *Summary of Study Design*

Study	Scenario	Design	Measures
Study 1	3 x Fire, train, factory	2 Outcome (positive, negative) x 3 Distal event (positively motivated action, negatively motivated action, physical event)	<ul style="list-style-type: none"> • Cause • Blame • Unitisation of Chain • Voluntariness (manipulation check)
Study 2	3 x Same as Study 1	Same as Study 1	<ul style="list-style-type: none"> • Cause (same as Study 1) • Praise/blame (bipolar blame measure) • Unitisation of Chain (same as Study 1) • Responsible
Study 3	3 x Same as Study 1	Same as Study 1	<ul style="list-style-type: none"> • Cause (same as Study 1) • Praise/blame (revised Study 2) • Responsible (same as Study 2) • Punitiveness • Probability
Study 4	5 x Same as Study 1 Plus avalanche and flooding	2 Outcome (positive, negative) x 4 Distal event (positive action, negative action, abnormal physical, normal physical) x 2 Proximal event (abnormal physical, normal physical)	<ul style="list-style-type: none"> • Same as Study 3 • Abnormality
Study 5	7 x Fire, train, factory, avalanche, car, casino and soldier.	2 Outcome (positive, negative) x 3 Distal event (positive action, negative action, abnormal physical)	<ul style="list-style-type: none"> • Cause (same as Study 1) • Praise/blame (same as Study 3) • Probability (same as Study 3)
Study 6	7 x Same as Study 5	2 Outcome (positive, negative) x 3 Distal motive (positive action, negative action, unstated motive action)	<ul style="list-style-type: none"> • Same as Study 5 • Intent (2 items: belief & desire) • Intentionality • Foresight (2 items: foresight, predictability) • Unitisation of Chain (revised Study 2)