Running Head: Manipulating Morality

Manipulating morality:
Third-party intentions alter moral judgments by changing causal reasoning

Jonathan Phillips[1] & Alex Shaw[2]

[1]Yale University [2]University of Chicago

Word count: 8,472
Figures: 9
Tables: 2

Corresponding author:
Jonathan Phillips
Department of Psychology
Yale University
P.O. Box 208205
New Haven, CT 06520
jonathan.phillips@yale.edu
(203) 654-9240

Abstract

The present studies investigate how the intentions of third parties influence judgments of moral responsibility for other agents who commit immoral acts. Using cases in which an agent acts under some situational constraint brought about by a third party, we ask whether the agent is blamed less for the immoral act when the third party intended for that act to occur. Study 1 demonstrates that third-party intentions do influence judgments of blame. Study 2 finds that third-party intentions only influence moral judgments when the agent's actions precisely match the third party's intention. Study 3 shows that this effect arises from changes in participants' causal perception that the third party was controlling the agent. Studies 4 & 5 respectively show that the effect cannot be explained by changes in the distribution of blame or perceived differences in situational constraint faced by the agent.

Keywords: Morality; Intention; Causation; Manipulation; Causal chains

1. Introduction

Recent research in moral psychology has led to significant discoveries about the factors that influence people's moral judgments. When people make moral judgments about an agent, they consider facts about what that agent *caused* (Cushman & Young, 2011), what that agent *believed* and *desired* (Cushman, 2008; Guglielmo & Malle 2010b; Pizarro, Uhlmann & Salovey, 2003; Shaver, 1985; Young & Saxe, 2008), what kind of *character* the agent has (Pizarro & Tannenbaum, 2011), what the agent was obligated to do (Hamilton, 1978), what sort of *control* the agent had over the outcome (Woolfolk, Doris & Darley, 2006; Weiner, 1995), and others as well. Yet, what these different research programs have in common is that they all investigate factors that are straightforwardly *about the agent* who is doing the immoral action.

While the reasons for this emphasis are obvious, there are also important instances in which people's moral judgments seem to depend on facts about other, third-party agents. Consider an agent who does something immoral after being manipulated. In such cases, we typically don't regard those agents to be fully morally responsible for their actions (Sripada, 2012). Moreover, this judgment of responsibility seems to depend on facts about what the other third-party agent did: the more the third party constrained the person's actions, the less responsible we find the agent to be. However, some have questioned whether third party *intentions*, not just the situational constraint, might influence people's judgment of responsibility. This issue has arisen, for example, in the philosophical discussion of manipulation. Derk Pereboom (2001) famously reflected on several relevant cases, including one in which a manipulator changes the agent's environment in such a way that it causes the agent to murder someone, and another in which the environment merely happens to be set up such that it causes the agent to murder that same person. While the philosophical discussion of manipulation is

nuanced in important ways (see, e.g., Mele, 1995; 2006; Rosen, 2002), one central argument has been that given that the actual situation the agent faced was the same in both cases, the manipulator's intentions are simply not relevant to questions of whether the manipulated agent is morally responsible (Greene & Cohen, 2004).

The philosophical debate concerning the relevance of moral responsibility to manipulation has certainly not been resolved (see, e.g., Fisher, 2004; McKenna, 2008; Pereboom, 2008), but it does raise important questions for moral psychologists. Are facts about third-party agents relevant to the way in which we ordinarily make moral judgments about other agents? And, if they are, why would facts about third-party agents influence the moral judgments we make about the agent who actually commits the immoral action? We take up these questions here, employing cases of manipulation in which an agent acts under some situational constraint brought about by a third party. We first find evidence that the intentions of third-party agents do influence judgments of moral responsibility: agents exposed to the exact same situational constraint receive less blame when that constraint was intentionally created by a third party. Additionally, we demonstrate that these effects are accounted for by the way that third-party intentions alter participants' causal reasoning about the agents involved. That is, the third party is perceived as directly causing the manipulated agent to act when the third party created the situational constraint with the intention of prompting the agent to do something immoral.

1.1 *Moral judgment and intention*

Agents' intentions play an indisputably important role in ordinary moral judgments (for a recent review, see Waldmann, Nagel & Weigmann, 2012). This close connection has long been a point of discussion in philosophy, psychology and law (Kenny, 1973; Williams, 1993; see also, Aristotle, trans. 2002; Hume, 1751/1998). The standard view of the connection between intention

and moral judgment has been that agents should be blamed and punished more for outcomes that were intended than outcomes that were unintended (Borg, Hynes, Van Horn, Grafton & Sinnott-Armstrong, 2006; Piaget, 1965; Young, Bechara, Tranel, Damasio, Hauser & Damasio, 2010).

However, recent research has suggested that the relationship between intentions and moral judgment may not be quite so simple. For example, not only do judgments of intention influence moral judgment, but moral judgment can influence judgments of intention as well (Knobe, 2003; Leslie, Knobe & Cohen, 2006; Phelan & Sarkissian, 2008; see also, Alicke, 2008; Guglielmo & Malle 2010a; Nadelhoffer, 2006; Uttich & Lombrozo, 2010). While such results highlight the complexity of the relationship between intention and moral responsibility, a related set of research has also helped to clarify that relationship by demonstrating the way in which some puzzling aspects of moral judgment can be explained by appealing directly to how and when we represent an agent's intentions (for a review, see Young & Tsoi, 2013). To take one example, intentions matter more for some types of moral violations than others: moral judgments about violations that involve disgust (incest, food taboos, etc.) are less sensitive to the intentions behind the action (e.g., if one was not aware that one's sexual partner was, in fact, a sibling, and did not intend to commit incest) than moral judgments about harm (killing, stealing, etc.) (Young & Saxe, 2011). This difference can be explained by the differential role intentions have in affective disgust-responses and affective anger-responses respectively (Young & Tsoi, 2013). The present research **is similar in that it appeals to nonmoral facts** about how we reason about other agents' intentions to explain otherwise puzzling patterns of moral judgment (Cushman & Young, 2011; Young & Saxe, 2011). Unlike previous research, though, it focuses on the underexplored connection between third-party intentions and morality.

1.2 *Moral judgment and multiple agents*

While we are not aware of previous research investigating third-party intentions in cases of manipulation, earlier research (Fincham & Schultz, 1981) did consider related cases involving causal chains with multiple agents contributing to a negative outcome. This research found that when a proximal agent (i.e., the agent who directly caused the negative outcome) voluntarily committed an immoral action that directly brought about the negative outcome, participants placed less blame on more distal agents (i.e., those agents who contributed but not directly cause the outcome) for the negative outcome that arose. In these cases, the presence of a proximal agent who is clearly morally responsible for the outcome seems to render more distal agents less responsible.

Similarly, more recent research found that the mere presence of an intermediate agent between an outcome and a more distal agent can impact attributions of moral responsibility to the distal agent (Bartling & Fischbacher, 2012; Coffman, 2011; Hamman, Loewenstein, & Weber, 2010; Paharia, Kassam, Greene, & Bazerman, 2009), e.g. people blame someone less for being selfish when they use an intermediary agent to achieve their selfish end. This effect persisted even when the distal agent knew what the intermediary would do and involved them in the causal chain for that reason (Coffman, 2011, Paharia et al. 2009).

While related, the present research deviates from these previous studies in an important way. Focusing specifically on the *intentions* of the *distal* agent, we investigate whether these mental states can change the moral responsibility of a proximal agent who is completely unaware of the distal agent's mental states. Accordingly, the phenomenon under investigation is comparatively more puzzling, as it involves the mental states of a distal agent (who does not actually cause the outcome) changing the moral responsibility of the proximal agent who actually does cause the outcome.

1.3 Moral judgment and causation

Like intention, causation has long been acknowledged as central to moral responsibility in that we hold agents who cause bad outcomes to be morally responsible for them, but not agents who don't actually cause them (see, e.g., Hart & Honoré, 1959). Moreover, this close connection between causation and morality has been used by researchers to show that idiosyncratic patterns of causal reasoning are often mirrored in related moral judgments (Mikhail, 2011; Waldmann & Dieterich, 2007; Weigmann & Waldmann, 2014). To take one example, Cushman (2008) demonstrated that people assign *more* punishment to failed attempts at causing a harm when no harm occurs, than when a harm actually does occur but is brought about through an independent causal mechanism. This effect of 'blame blocking' seems to arise directly from differences in ordinary causal reasoning about the two cases. Or to take another example, the severity of people's moral judgments depends in part on how 'deviant' the causal chain is that leads to the outcome, regardless of the badness of the outcome and what the agent intended (Pizarro, Uhlmann & Bloom, 2003).

The present study contributes to this research by demonstrating another way in which causal reasoning and moral responsibility are connected. However, rather than focusing on the causal reasoning about one particular agent's contribution to the outcome, we explore the causal relationship between multiple agents who all contribute to a negative outcome. More specifically, we consider the role of *intentions* in changing this causal relationship.

1.4 Present research

Here, we explore whether third-party intentions influence moral judgments of the proximal agent when that agent is being manipulated through the introduction of some situational constraint. In five studies, we demonstrate that judgments of moral responsibility are sensitive to the intentions

of third parties and provides evidence that third-party intentions influence judgments of moral responsibility by changing participants' causal reasoning.

2. Study 1

Study 1 specifically asked whether judgments about the moral responsibility of a manipulated agent (the proximal agent) are influenced by the intentions of a third-party manipulator (the distal agent). In this and the following studies, participants read vignettes in which a distal agent brings about some situational constraint which induces the proximal agent to commit an immoral act. Importantly, in all cases the agent is never aware of the distal agent's intentions.

We predicted that, even holding fixed both the proximal agent's situational constraint and all of the distal agent's actions, the distal agent's mental states alone would influence participants' judgments of blame. More specifically, we predicted that participants would blame the proximal agent less when the distal agent created the situation with the intention of inducing the proximal agent to commit an immoral act.

2.1 Method

2.1.1 Participants

Participants were 78 adults (34 females; Age: $M = 30.14$, $SD = 8.65$) on Amazon's Mechanical Turk (Buhrmester, Kwang & Gosling, 2011; Gosling, Vazire, Srivastava & John, 2004). In all studies, participation was restricted to participants located in the United States with above a 95% approval rating and were paid a small amount of money (less than $0.50) for participating. No participants were excluded from any analysis in this paper as we believe that, when possible, the best practice is to practice is to collect larger samples and thereby avoid having to determine arbitrary exclusion criteria.

2.1.2 Procedure

Participants were randomly assigned to either the ‘intentional’ or the ‘accidental’ condition.[1] All participants read a short vignette that described a group of industrial workers who raided a small village, stole food, and murdered innocent people. In both cases, the industrial workers were prompted to raid the small village because the government (the distal agent) caused a severe food shortage. In the intentional condition, the government intended for the food shortage to cause the workers to attack the small village. As these vignettes are used in several subsequent studies, we present them in full:

> In the 1950s, the government of a small Eastern European country plotted to secretly start a war, using industrial workers, and get revenge on a neighboring country. For the first part of their plan, the government intentionally destroyed farm machinery and set fire to several food stores on purpose. As a result, there was a serious lack of food in the country. Soon the people living in the city couldn't get enough food to feed themselves. The whole city shut down, crime skyrocketed and a small but violent uprising broke out.
>
> The government knew their plan was working perfectly. Right at that time, a group of industrial workers heard on the government news channel that a neighboring village had a surplus of food. After hearing the news, the group of industrial workers raided the small village on the country's border, stealing food from the farmers and killing innocent people. The government had known this would happen all along and felt great about their successful plan.

[1] All stimuli, analyses and data can be found at: https://github.com/phillipsjs/Manipulation

In the accidental condition, however, the government accidentally caused the food shortage (differences in italics):

> In the 1950s, the government of a small Eastern European country planned to start a new *economic program*, using industrial workers, *to increase the country's wealth*. *While it wasn't* part of their plan, the government *accidentally* destroyed farm machinery and set fire to several food stores *by mistake*. As a result, there was a serious lack of food the country. Soon the people living in the city couldn't get enough food to feed themselves. The whole city shut down, crime skyrocketed and a small but violent uprising broke out.
>
> The government knew their plan *wasn't working at all*. Right at that time, a group of industrial workers heard on the government news channel that a neighboring village had a surplus of food. After hearing the news, the group of industrial workers raided the small village on the country's border, stealing food from the farmers and killing innocent people. The government had *not* known this would happen and felt *terrible* about their *unsuccessful* plan.

After reading the vignettes, participants were asked whether they agreed or disagreed with the statement "The workers should be blamed for the attack on the village." Secondly, they were asked whether they agreed or disagreed with the statement "The government should be blamed for the attack on the village."

2.2 Results

Blame ratings were analyzed using a 2 (Condition: Intentional vs. Accidental) x 2 (Agent: Manipulator vs. Manipulee) mixed ANOVA. This revealed no effect of condition, $F<1$, a marginal effect of Agent, $F(1,150)=3.53$, $p=.062$, and a Condition x Agent interaction,

F(1,150)=8.05, *p*=.005. The interaction was decomposed with planned comparison which revealed that blame ratings for the government (the distal agent) were lower in the accidental condition (*M* = 5.05, *SD* = 1.28) than in those in the intentional condition (*M* = 6.27, *SD* = 0.84), *t*(69.50) = -5.02, *p* < .001, Cohen's *d* = 1.12. (Fig. 1). More importantly, blame attributions for the industrial workers (the proximal agent) were significantly greater in the accidental condition (*M* = 5.32, *SD* = 1.51) than those in the intentional condition (*M* = 4.35, *SD* = 1.62), *t*(73.73) = 2.71, *p* = .008, Cohen's *d* = 0.62.

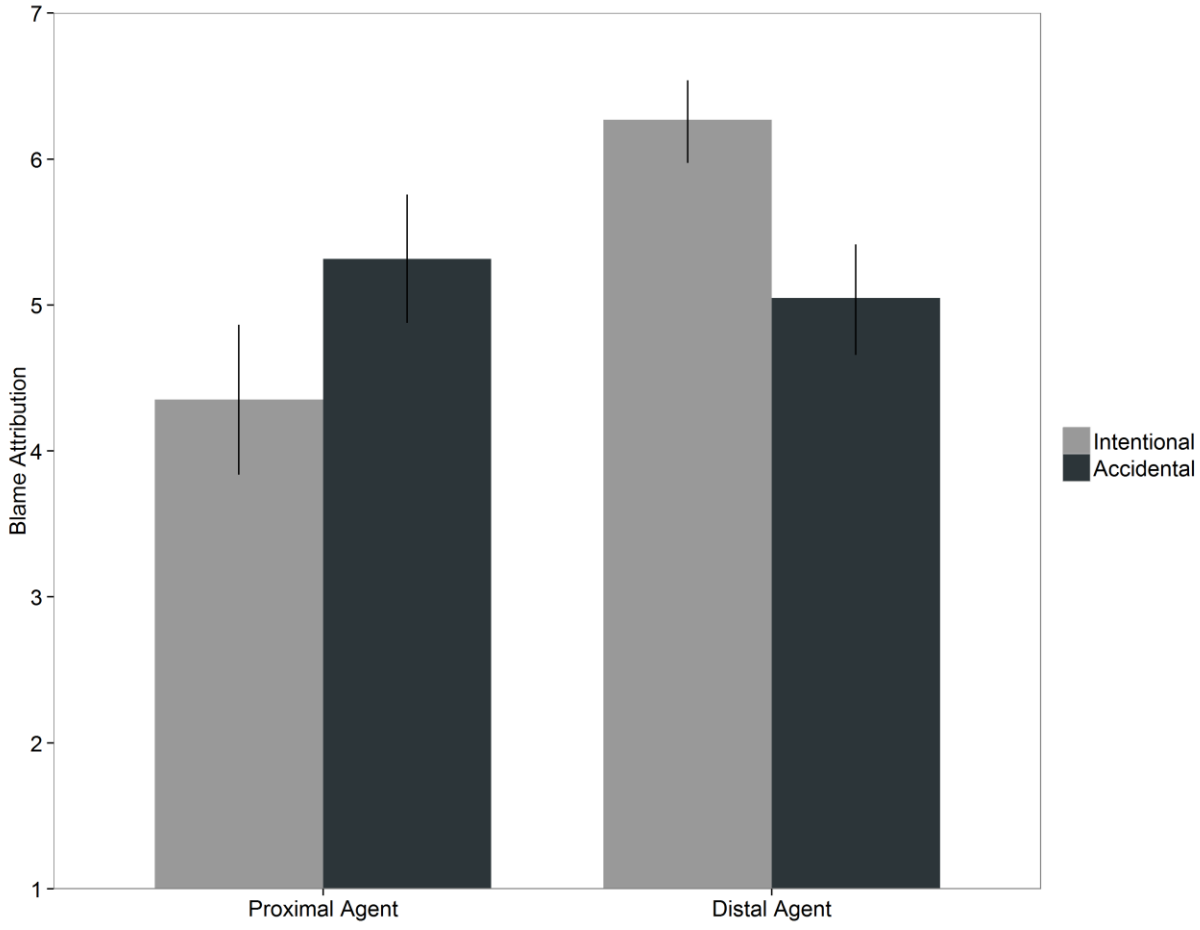

Figure 1. Judgments of Blame for the Distal and Proximal Agents in Intentional and Accidental Conditions. Error Bars Indicate Bootstrapped 95% CIs.

2.3 Discussion

Third-party intentions altered participants' judgments of blame. This effect occurred even though participants were not given any reason to believe that the proximal agent was aware of the distal agent's intentions, suggesting that from the workers' perspective the two situations would have been identical and thus should have provided the same situational constraint. Despite this similarity, participants' judgments of blame for the proximal agent changed depending on the psychological states of the distal agent. While this study provides initial evidence that third-party intentions exert an influence on moral responsibility judgments, an even better test of this hypothesis would consider not the presence or absence of intention per se (which might reasonably influence the distal agent's actions) but the precise match between the proximal agent's action and the specific *content* of the distal agent's intention.

3. Study 2

To test whether moral judgments of the proximal agent's actions really were being influenced by the intentions of the distal agent, this study investigates whether the effect arises in a more minimal pair. Specifically, this study holds fixed both the intentions and the actions of the distal agent, and changes only the extent to which the proximal agent's actions match the content of the distal agent's intentions. In the Intention-consistent condition, the proximal agent's action is exactly consistent with the distal agent's intentions, while in the Intention-deviant condition the proximal agent's action deviates slightly from the specific content of the distal agent's intentions. As before, the proximal agent has no knowledge of the distal agent's intentions.

3.1 Method

3.1.1 Participants

Participants were 334 users of Amazon's mechanical Turk website (128 females; Age: M = 32.14, SD = 12.08). A larger number of participants were recruited as the difference between the two conditions was necessarily very minimal given the aim of the study. The procedures for Study 2 were identical to the previous study except for the changes in the vignettes read by participants.

3.1.2 Procedure

All participants read one of two vignettes, both of which began identically to the vignette from Study 1, but ended in slightly different ways. In the intention-consistent condition, the workers attacked the village that the government intended for them to attack, while in the intention-deviant condition, the workers attacked a different village instead:

> The government knew their plan was working perfectly. Right at that time, a group of industrial workers heard through the government news channel that the Shaki village had a surplus of food. After hearing the news, the group of industrial workers raided [the Shaki village/a village on the opposite side of the small country called the Nobi village], stealing food from the farmers in that village and killing innocent people. The government [had / had not] known this would happen all along and it [was exactly what they planned / wasn't part of their plan]. They were [pleased that everything worked just as they knew it would. / dismayed that nothing worked the way they thought it would.]

Once again, participants were first asked whether they agreed or disagreed with the statement, "The workers should be blamed for the attack on the village." Secondly, they were asked

whether they agreed or disagree with the statement, “The government should be blamed for the attack on the village.”

3.2 Results

Participants’ blame ratings were analyzed using a 2 (Condition: Consistent vs. Deviant) x 2 (Agent: Manipulator vs. Manipulee) mixed ANOVA. This revealed a marginal effect of Condition, $F(1,662)=3.11$, $p=.078$, a main effect of Agent, $F(1,662)=63.07$, $p<.001$, and critically, a Condition x Agent interaction effect, $F(1,662)=5.86$, $p=.016$. The interaction was decomposed with planned comparison which revealed that blame attributions for the proximal agent (the industrial workers) were significantly greater in the Deviant condition ($M = 4.27$, $SD = 1.74$) than those in the Intentional condition ($M = 3.89$, $SD = 1.70$), $t(331.23) = 2.05$, $p = .041$, Cohen’s $d = 0.22$. However in contrast to the previous study, blame ratings for the distal agent (the government) did not differ (Deviant: $M = 6.01$, $SD = 1.11$; Consistent: $M = 6.05$, $SD = 1.26$), $t(328.32) = -0.32$, $p = .752$. (Fig. 2)

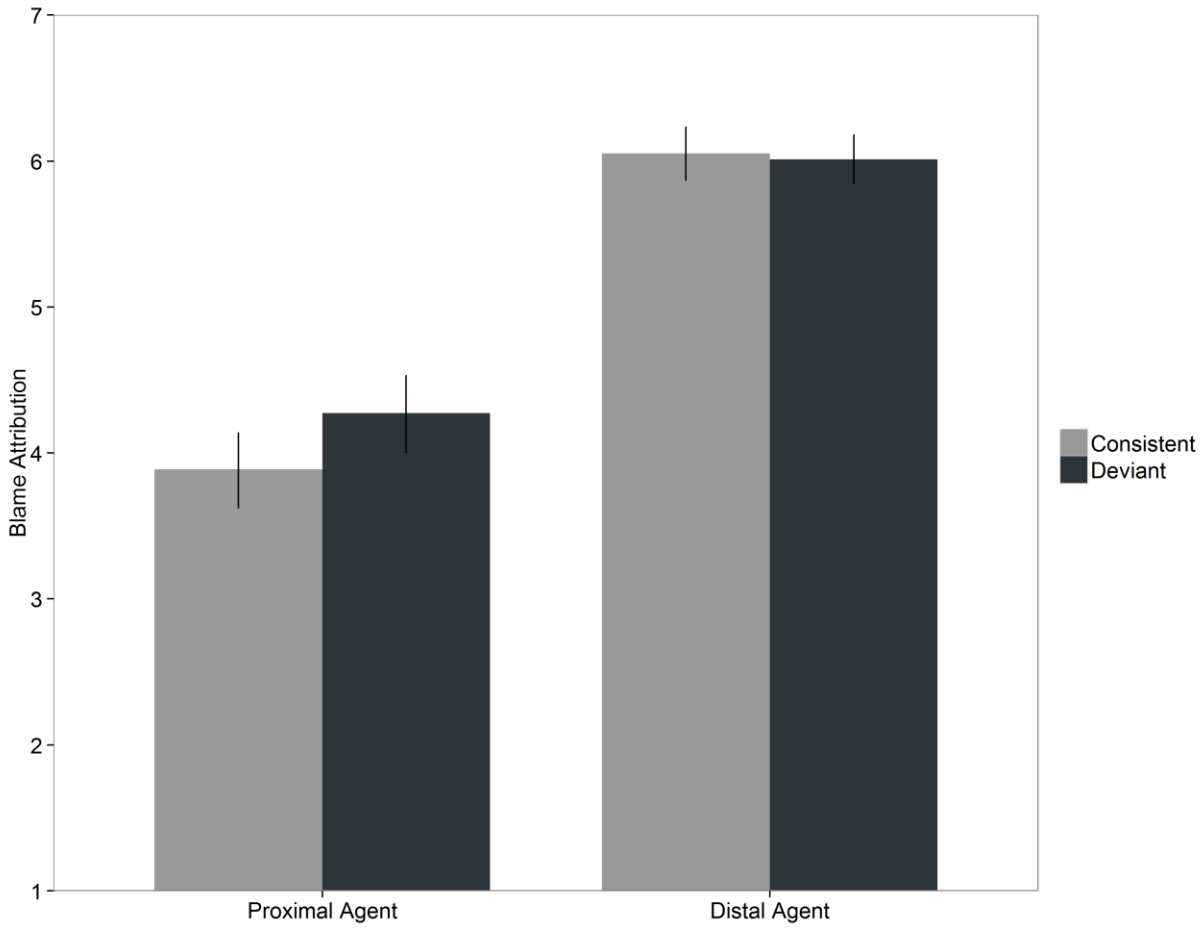

Figure 2. Judgments of Distal and Proximal Agents in the Intention-consistent Condition (dark bars) and Intention-deviant Condition (light bars). Error Bars Indicate Bootstrapped 95% CIs.

3.3 Discussion

Participants blamed the proximal agent less for killing and stealing when the action conformed to the third party's intention than when it deviated even slightly from it. This effect occurred even though the distal agent intended to manipulate the proximal agent into attacking a village in both conditions. This effect, therefore, seems to be highly specific to the intentions of the distal agent.

Moreover, as the effect occurred in a context in which blame ratings for the distal agent (the government) did not significantly differ, it is unlikely that the effect could be explained by

differences in the distribution of blame between the distal and proximal agents, or by an implicit comparison between the two agents (Study 4 provides additional evidence for this). If this explanation does not account for our findings, then what does explain them?

4. Study 3

One explanation of the effect presented in Studies 1 and 2 is that the difference in participants' moral judgments in the two scenarios is driven by differences in their causal reasoning. This approach falls roughly in line with a growing body of research that suggests that moral judgments are often derived from non-moral psychological representations (Cushman, 2008; Cushman & Young, 2011; Pizarro, Uhlmann & Bloom, 2003). Suggestively, previous research in causal cognition has provided evidence that whether or not the distal agent foresaw or intended the eventual outcome can impact whether the distal agent is perceived as causing that eventual outcome (Fincham & Schultz, 1981; Lombrozo, 2010). Accordingly, it is possible that results from previous research in causal cognition can be employed to help explain the effect of third-party intentions on moral judgment.

However, a slight extension of the proposal from previous research is required to account for these effects: while previous research has demonstrated that the distal agent's intentions can influence whether the distal agent is seen as causing the *outcome*, our proposal is that this previous work can be extended to suggest that the distal agent's intentions also change whether the proximal agent's actions are seen as caused by the distal agent. If this is true, then the reason the proximal agent was judged to be less morally responsible would become clear: the proximal agent was perceived as having been made to do the immoral action.

To return to the example in Studies 1 and 2, when the government intends for the workers to attack a particular village and they do so, participants could have been led to perceive the government as causing the workers' actions. By contrast, when the distal agent does not intend for the proximal agent to commit an immoral act (or when the proximal agent's actions do not exactly match the distal agent's intention), participants should not perceive the proximal agent's actions as being controlled in the same way. Instead, the proximal agent should be perceived as more freely choosing their actions, and be blamed accordingly. We test this proposal by using a meditational analysis to assess whether third-party intentions influence proximal-agent blame by changing the causal judgments of the distal agent.

4.1 Method

4.1.1 Participants

Participants were 123 users of Amazon's mechanical Turk website (70 females). The procedure for Study 3 was identical to Study 1 except for the questions participants were asked.

4.1.2 Procedure

Participants were randomly assigned to read either the intentional or the accidental version of the vignettes used in Study 1. After reading the vignettes, participants were asked to give agreement ratings with a series of statements.

To test blame for the proximal agent, participants were once again asked to rate their agreement with the measure used in previous studies.

Proximal Agent Blame: "The workers should be blamed for the attack on the village."

Secondly, to test the prediction of third party causation, participants also rated their agreement with three statements presented in counterbalanced order:

Distal Agent Made: "The government made the workers attack the village."

Distal Agent Caused: "The government caused the workers to attack the village."

Because of Distal Agent: "The workers attacked the village because of the government."

4.2 Results

4.2.1 Primary Analyses

Once again, participants blamed the proximal agents less when the distal agent intended for them to do the immoral action (M = 5.06, SD = 1.52) than when the distal agent did not (M = 4.21, SD = 1.58), $t(118.96) = 3.07$, $p = .003$. The three causation items were combined into a single scale of distal-agent causation ($\alpha = .76$). Participants also more agreed that the distal agent caused the proximal agent's action in the intentional condition (M = 5.02, SD = 1.02) than in the accidental condition (M = 3.98, SD = 1.25), $t(122.58) = 5.14$, $p < .001$. (Fig. 3)

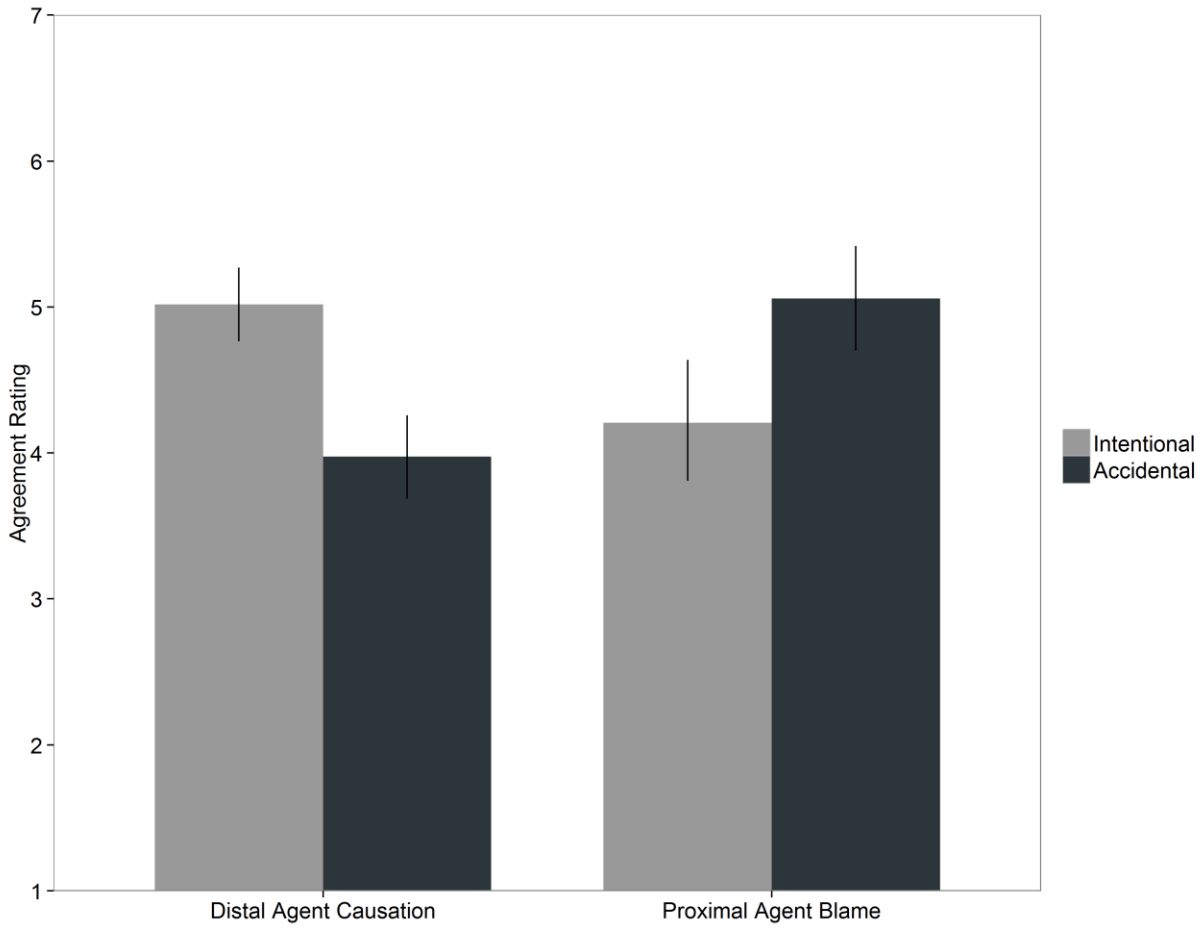

Figure 3. Mean Ratings of Proximal Agent Blame and Distal Agent Causation in Accidental vs. Intentional Conditions. Error Bars Indicate Bootstrapped 95% CIs.

4.2.2 Mediational Analyses

A bootstrapped mediation analysis (Preacher & Hayes, 2008) was used to test whether the distal agent causation scale mediated the effect of condition on judgments of proximal agent blame (Fig. 4). The indirect effect coefficient for distal agent causation was -.278, and the bias corrected 95% confidence interval ranged from -.608 to -.007. As this confidence interval does not include zero, distal agent causation was indeed a significant mediator. The total effect condition on proximal agent blame decreased from a total effect of -.853, p = .003 to a direct

effect of -.582, p = .055. Moreover, the alternative mediation model in which judgments of proximal blame mediate the effect of distal agent intentions on distal agent causation was not significant, 95% CI [-.002, .328].

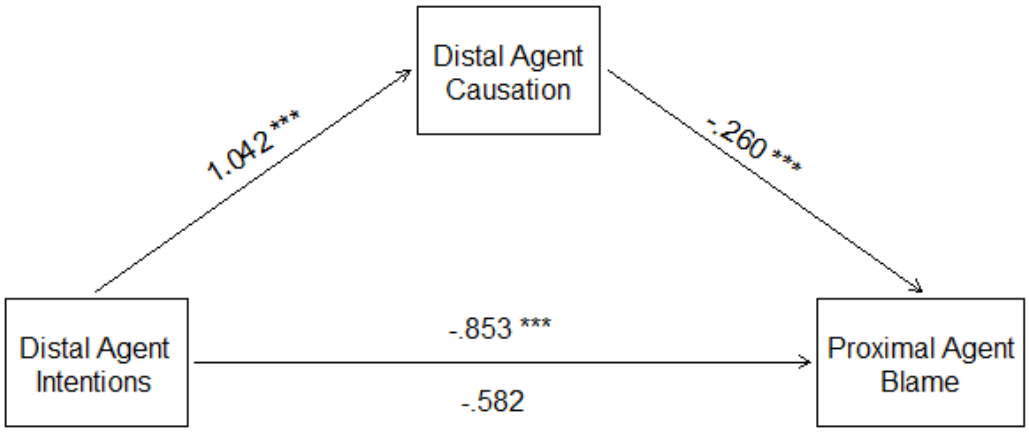

Figure 4. Mediation Analysis with Distal Agent Causation Mediating the Effect of Distal Agent Intentions on Proximal Agent Blame. * p < .05. ** p < .01. *** p < .001.

4.3 Discussion

These analyses provide evidence that the impact of third-party intentions on proximal agent blame is caused by changes in participants' causal reasoning about the events that occurred. More specifically, this study suggests that the distal agent's intentions changed the extent to which participants perceived the proximal agent's action as being caused by the distal agent, and the less they saw their actions as caused, the more they held the proximal agent morally responsible.

This finding extends previous research on the role of intentions in causal cognition (Lombrozo, 2010; Fincham & Schultz, 1981) and adds to the growing literature which suggests that moral judgments are often derived from non-moral psychological representations (Cushman, 2008; Cushman & Young, 2011; Pizarro et al, 2003). Still, the mechanism investigated in this study is not the only plausible one by which the distal agent's intentions could change attributions of blame for the proximal agent. We investigate a second potential mechanism in Study 4 and compare it to this causal mechanism.

5. Study 4a

Another plausible explanation, previously proposed in research on the effect intermediary agents on moral judgment, would be that the present effect arose because of changes in how blame was distributed between the two agents (Hamman, et al., 2010; Bartling & Fischbacher, 2012). Reframing this suggestion within the present study, the basic proposal would be that the proximal agent was blamed less when the distal agent intended the bad outcome because the distal agent absorbed more blame for that outcome. We refer to this basic mechanism as the *conservation of moral responsibility*, as it suggests that when multiple agents are responsible for an outcome, the more we hold one agent morally responsible, the less we end up holding the other responsible. This explanation is a plausible alternative that merits investigation.

Given the initial plausibility of both a conservation of blame mechanism and the evidence for the causal mechanism from Study 3, we now turn to a test between them. To test these two proposed explanations against each other, we introduce three new measures of moral responsibility for the distal agent, and using multiple mediation, examine whether distal agent causation or distal agent moral responsibility is the better mediator of the effect of distal agent's

intentions. If the distal agent's intentions directly change the distribution of blame, then judgments of moral responsibility for the distal agent should be a better mediator of blame judgments for the proximal agent. By contrast, if the distal agent's intentions indirectly change blame for the proximal agent by changing whether participants perceive the proximal agent's actions as caused by the distal agent, then distal agent causation should be a better mediator.

In addition to adding three new measures of distal agent moral responsibility, Study 4 introduces three new pairs of scenarios which again differ only in the mental states of a third-party. This provides at least two benefits. The first is that it ensures that the effect of third-party intentions on proximal agent blame generalizes beyond the scenario used in previous studies. The second is that it ensures that any mediation patterns must generalize across many highly different cases of manipulation.

5.1 Method

189 participants were recruited from Amazon's Mechanical Turk website (66 females; Age: $M = 30.77$, $SD = 10.15$). Participants were randomly assigned to read either the Intentional or Accidental condition of one of the four different scenario-pairs that varied only the intentions of the distal agent. After reading one vignette, participants indicated their agreement with an attribution of blame for the proximal agent.

Proximal Agent Blame: [The proximal agent] should be blamed for [the outcome].

Afterward, participants rated their agreement with three attributions of causation to the distal agent causation and three attributions of moral responsibility to the distal agent in random order:

Distal Agent Causation:

- [Proximal Agent] did [immoral action] *because* of [Distal Agent].

- [Distal Agent] *caused* [Proximal Agent] to do [immoral action].
- [Distal Agent] *made* [Proximal Agent] do [immoral action].

Distal Agent Moral Responsibility:

- [Distal Agent] is *bad*.
- [Distal Agent] should be *blamed*.
- [Distal Agent] acted *wrongly*.

5.2 Results and Discussion

5.2.1 Primary Analyses

Participants' judgments of blame for the proximal agent were analyzed with a 2 (Condition: Intentional versus Accidental) x 4 (Scenario) ANOVA. We observed a main effect of distal agent intentions $F(1, 397) = 17.60$, $p < .001$ $\eta_p^2 = 0.042$, such that participants blamed the agent less in the Intentional condition ($M = 4.48$, $SD = 1.85$) than in the Accidental condition ($M = 5.21$, $SD = 1.74$). We also observed a main effect of Scenario, $F(3, 397) = 10.64$, $p < .001$, $\eta_p^2 = 0.074$, suggesting that participants found the actions in some scenarios worse overall, but no Condition x Scenario interaction, $F < 1$. The higher attributions of blame for the proximal agent in the Accidental versus Intentional conditions across this new set of scenarios serves both as a replication and as an extension of the previous studies.

Participants' responses for the three items measuring distal agent causation were combined into a single scale ($\alpha = .85$). The distal agent causal scale was then subjected to a 2 (Condition: Intentional versus Accidental) x 4 (Scenario) ANOVA. We observed a main effect of distal agent intentions $F(1, 185) = 72.13$, $p < .001$, $\eta_p^2 = 0.279$, such that participants perceived the distal agent as more causing the proximal agent's action in the Intentional condition ($M =$

5.45, *SD* = 1.35) than in the Accidental condition (*M* = 3.69, *SD* = 1.56). There was no main effect of Scenario, *F*(3, 185) = 1.78, *p* = .153 and no Condition x Scenario interaction effect, *F*(3,185) = 1.53, *p* = .208.

Participants' responses for the three items measuring distal agent moral responsibility were also combined into a single scale (α = .89). The distal agent moral responsibility scale was subjected to a 2 (Condition: Intentional versus Accidental) x 4 (Scenario) ANOVA. We observed a main effect of Condition *F*(1, 185) = 294.13, *p* < .001 η_p^2 = 0.611, such that participants perceived the distal agent as more immoral in the Intentional condition (*M* = 6.33, *SD* = 0.92) than in the Accidental condition (*M* = 3.38, *SD* = 1.45). There was no main effect of Scenario, *F*(3, 185) = 1.09, *p* = .354 and no Condition x Scenario interaction effect, *F*(3,185) = 2.54 *p* = .058. (Fig. 5)

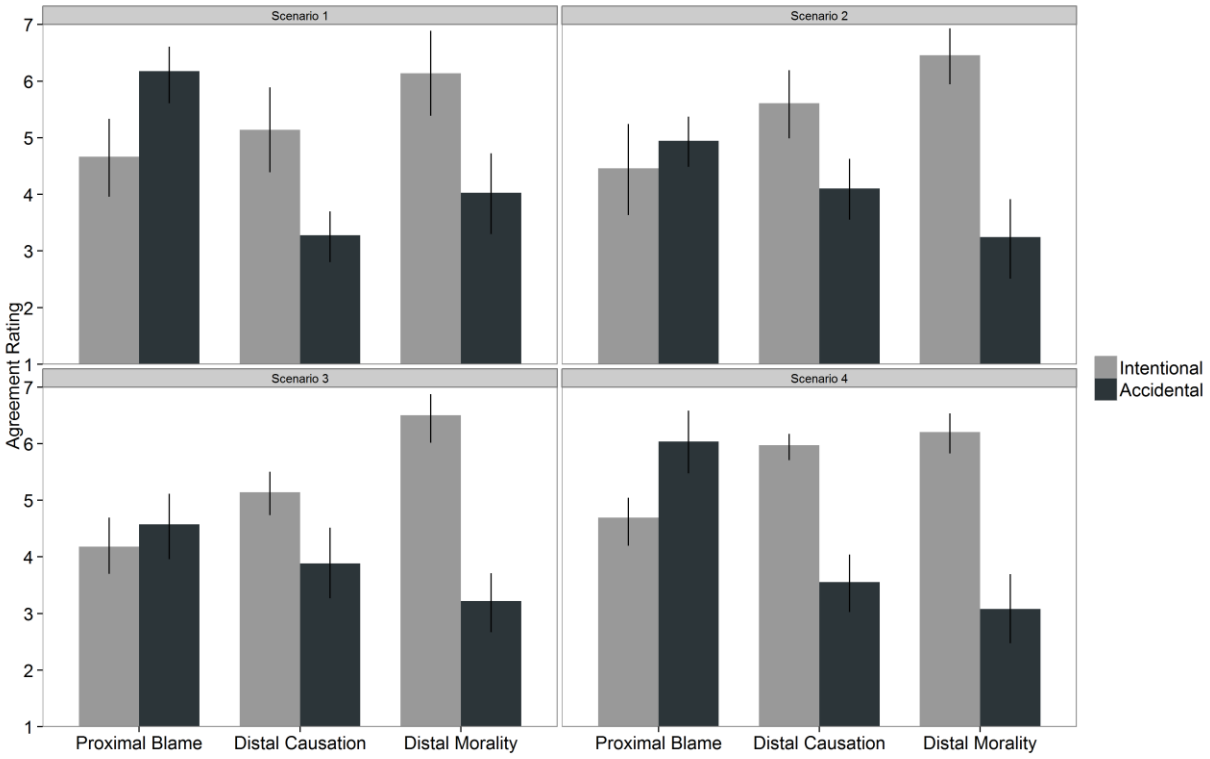

Figure 5. Mean Ratings of Agent Blame, Distal Agent Causation and Distal Agent Moral Responsibility in the Intentional vs. Accidental Conditions for All Four Scenarios. Error Bars Indicate Bootstrapped 95% CIs.

5.2.2 Mediational Analyses

To directly test which scale was a better mediator of the impact of the distal agent's intentions on blame for the proximal agent, we used a bootstrapped multiple mediation analysis (Preacher & Hayes, 2008), which allowed for the simultaneous calculation of both the indirect and direct effects of multiple mediators.

Both distal agent causation and distal agent moral responsibility scales were entered as mediators of the effect of third-party intentions on proximal agent blame. Moral responsibility judgments for the distal agent were not a significant mediator (bootstrapped indirect effect = .363, bias corrected 95% CI: [-.28, .96]. By contrast, distal agent causation was a highly significant mediator (bootstrapped indirect effect = -.800, bias corrected 95% CI: [-.1.21, -.45]. Moreover, in this mediation model, the effect of the distal agent's intentions on proximal agent blame decreased from a total effect of -.962, $p < .001$ to an direct effect of -.540, $p = .167$. (Fig. 6) These results both replicate the effect of third-party intentions on proximal agent blame, and provide further evidence that this effect is best explained by changes in participants' causal reasoning.

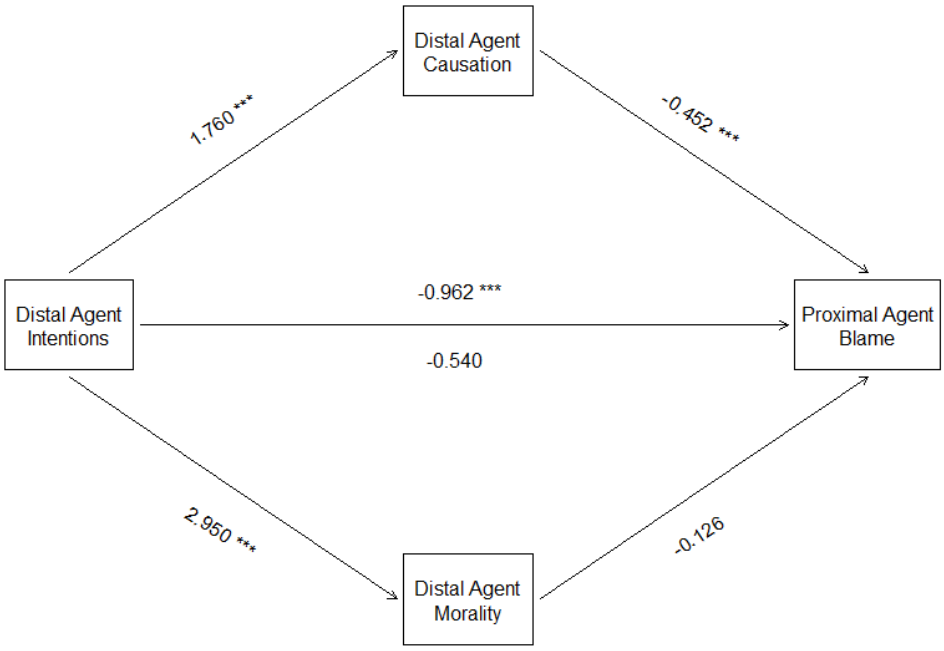

Figure 6. Multiple Mediation Analysis with Distal Agent Causation Mediating the Effect of Distal Agent Intentions on Proximal Agent Blame. *p < .05, **p <.01, ***p<.001.

One potential worry with this analysis is that the distal agent moral responsibility and causation scales were themselves significantly correlated, $r_{partial}(193) = .553$, $p < .05$. Accordingly, it is possible that we did not observe a significant mediation effect for the moral responsibility scale because this effect was artificially attenuated by the multicollinearity of the mediators (Preacher & Hayes, 2008). That is, the distal agent's moral responsibility may also partially mediate the reduction in blame to the proximal agent, but we were unable to observe this effect because of the high correlation between the two scales. To address this concern, an additional mediation analysis was conducted in which *only* the distal agent moral responsibility scale was included as a potential mediator. We once again found that the moral responsibility

scale did not significantly mediate the effect distal agent intentions on proximal agent moral responsibility, 95% CI: [-1.13, .08].

To further ensure that this mediation pattern was robust, we conducted a multiple mediation analysis for each causation-moral responsibility pair, along with the partial correlation between the two measures. These analyses revealed that for every causation-moral responsibility pair, regardless of the strength of the partial correlation, only the causal item was a significant meditator of the effect of distal agent intentions on proximal agent blame. (Table 1)

Finally, to ensure that this mediation pattern was replicable, a multiple mediation test is replicated in Study 4b with (1) a larger sample size and (2) using only the causal and moral pair that had the lowest partial correlation:

Distal Agent Causation: [The third-party] made [the proximal agent] do [the action].

Distal Agent Moral Responsibility: [The third-party] acted wrongly.

6. Study 4b

6.1 Method

Participants were 405 users of Amazon's Mechanical Turk website (128 females; Age: M = 29.85, SD = 10.40).

6.1.2 Procedure

All procedures for Study 4b were identical to Study 4a except that participants were asked only one distal agent causation and one distal agent moral responsibility item. After reading a randomly assigned vignette, participants completed the proximal agent blame measure and then the two mediator measures in counterbalanced order.

6.2 Results

6.2.1 Primary Analyses

Means for all three measures (*Proximal Agent Blame*, *Distal Agent Causation*, *Distal Agent Moral Responsibility*) are displayed in Figure 7. Each of these measures was analyzed with a 2 (Condition: Intentional versus Accidental) x 4 (Scenario) ANOVA. For each measure, we observed a main effect of Condition (p's < .001) and a main effect of Scenario (p's < .005). Additionally, we only observed an interaction effect for the *Distal Agent Moral Responsibility* measure (p = .008). (See Table 2 for full analyses and details.)

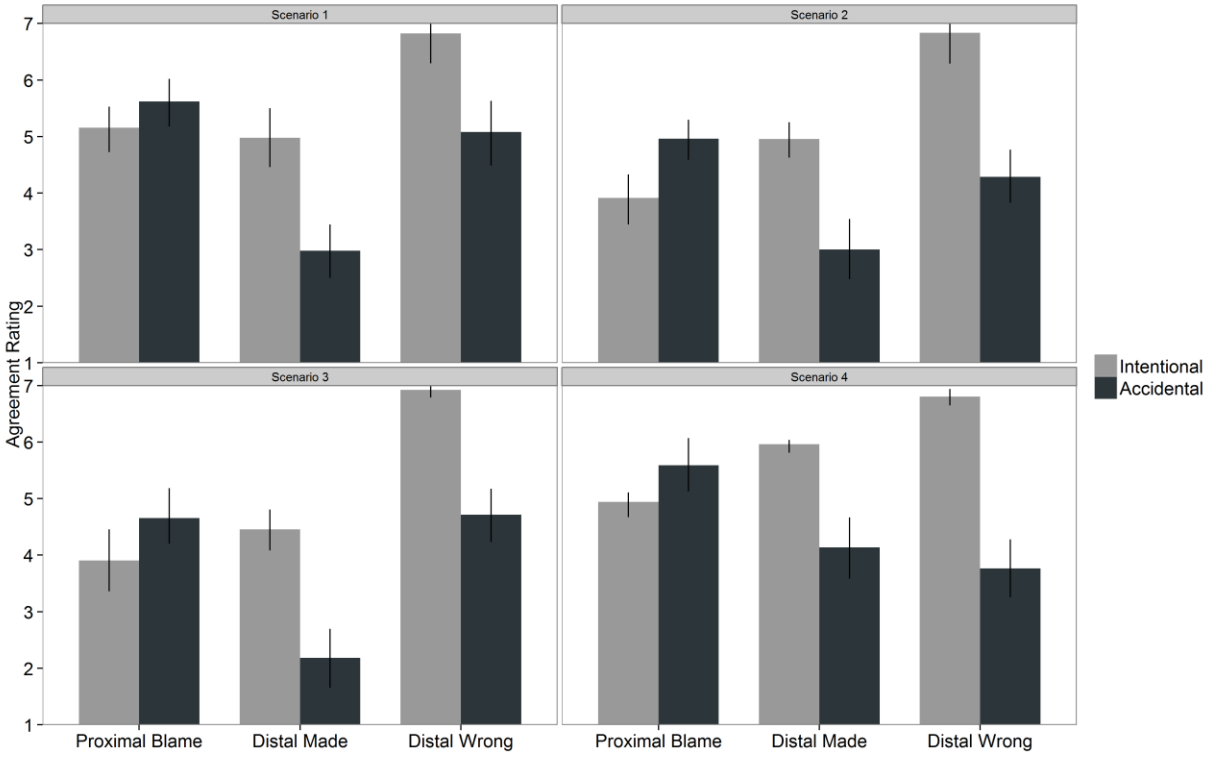

Figure 7. Mean Ratings of Agent Blame, Distal Agent Blame and Distal Agent Made in the Intentional vs. Accidental Conditions for All Four Scenarios. Error Bars Indicate Bootstrapped 95% CIs.

6.2.2 Mediational Analyses

A Variance Inflation Factor (VIF) analysis was conducted on the two causation and moral responsibility items to test for unacceptable rates of multicollinearity. This analysis revealed a VIF below 2, suggesting rates of multicollinearity far below an acceptable threshold (conservative treatments of multicollinearity suggest a VIF below 5, while more liberal treatments tolerate a VIF of up to 10 (Kutner, 2004)). These results suggest that a multiple mediation analysis may be appropriately conducted. Accordingly, distal agent moral responsibility and causation ratings were entered as mediators of the effect of third-party intentions on proximal agent blame. Moral responsibility judgments for the distal agent were again not a significant mediator (bootstrapped indirect effect =.017, bias corrected 95% CI: [-.312, .374]. By contrast, causal judgments of the distal agent were once again a highly significant mediator (bootstrapped indirect effect = -.493, bias corrected 95% CI: [-.742, -.284]. Moreover, the effect of the distal agent's intentions on blame for the proximal agent decreased from a total effect of -.725, $p < .001$ to a direct effect of -.256, $p = .280$, indicating that judgments of distal agent causation partially mediated the effect once again. (Fig. 8)

Figure 8. Multiple Mediation Analysis with Distal Agent Made Judgments Mediating the Effect of Distal Agent Intentions on Proximal Agent Blame. *p < .05, **p <.01, ***p<.001.

6.3 Discussion

We found that third-party intentions influence moral responsibility of proximal agents and provided evidence for the mechanism by which this effect occurs. While participants did find the distal agent to have acted much more wrongly when the distal agent intended to bring about a bad outcome, these judgments of moral responsibility did not bear a systematic relationship to judgments of blame for the proximal agent. By contrast, participants' judgments that the distal agent *made* the proximal agent do the immoral action did explain the reduction in blame for the proximal agent. The results of these mediation analyses strongly suggest that distal agents'

intentions affect moral judgments by changing causal reasoning. We further observed that third-party intentions influence moral responsibility judgments for an agent in a set of four highly different scenarios, demonstrating the robustness of the results.

7. Study 5

The previous studies provide evidence that third-party intentions affect participants' moral judgments by changing their causal reasoning. However, it remains unclear what exactly the relevant change in participants' causal reasoning was. One straightforward explanation for why participants could have perceived the distal as more causing the proximal agent's action in the intentional condition is that participants may have inferred that the distal agent created a higher level of situational constraint when acting intentionally. That is, although we described the situations identically in the two conditions, participants may have imagined that the situational constraint was actually stronger in the intentional condition. On this account, third-party intentions may change participants' moral judgments because the causal contribution of the distal agent through *the situational constraint* was greater when the distal agent acted intentionally. This explanation, while providing a novel role for third-party intentions, would situate the current studies as a demonstration of the well-established relationship between situational constraint and moral judgment (Woolfolk, Doris & Darley, 2006).

Alternatively, third-party intentions may also have a more direct impact on participants' causal reasoning in a way that does not involve changes in the situation constraint the proximal agent faced. On this second account, the match between the third party's intentions and the proximal agent's actions may lead participants to see the proximal agent as controlled by the third party from a distance, without changes in the situational constraint faced by the proximal

agent. In this final study, we investigate these two alternative causal mechanisms. Specifically, we add a direct measure of participants' judgments of whether the situational constraint caused the proximal agent to commit the immoral act. If the situational constraint explanation is correct, then participants should report that the situational constraint is more causal when it was created intentionally rather than accidentally. We additionally test whether this new measure mediates the impact of third-party intentions on moral judgment.

7.1 Method

Participants were 300 users of Amazon's Mechanical Turk website (100 females; Age: $M = 29.63$, $SD = 8.90$).

7.1.2 Procedure

All procedures for Study 5 were identical to Study 4b except that participants were either asked to rate their agreement with the *Distal Agent Causation* statement used in Study 4b or a question that asked about the influence of the situation on the proximal agent's action. In the case of the government and the industrial workers, for example, participants were asked to rate their agreement with the following statement:

> *Situational Causation*: The situation the workers were in (the destroyed farm machinery, the lack of food, etc.) caused the workers to attack the village.[2]

After reading a randomly assigned vignette, participants completed the proximal agent blame measure and then one of the two causal measures.

7.2 Results

[2] Full measures (with stimuli, analyses and data) are available at: https://github.com/phillipsjs/Manipulation

Participants' judgments of blame for the proximal agent were once again analyzed with a 2 (Condition: Intentional versus Accidental) x 4 (Scenario) ANOVA. We observed a main effect of condition $F(1, 292) = 21.83$, $p < .001$ $\eta_p^2 = 0.067$, such that participants blamed the agent less in the Intentional condition ($M = 4.20$, $SD = 1.97$) than in the Accidental condition ($M = 5.15$, $SD = 1.71$). We also observed a main effect of scenario, $F(3, 292) = 9.29$ $p < .001$ $\eta_p^2 = 0.087$, indicating that participants again found the actions in some scenarios worse overall, but no Condition x Scenario interaction, $p > .1$.

The analysis of the *Distal Agent Causation* measure again revealed a main effect of condition $F(1, 144) = 37.39$, $p < .001$ $\eta_p^2 = 0.178$, such that participants found the distal agent more strongly made the proximal agent do the immoral action in the Intentional condition ($M = 5.27$, $SD = 1.77$) than in the Accidental condition ($M = 3.51$, $SD = 1.89$). In addition, we also observed a marginal effect of scenario, $F(3, 144) = 2.21$, $p = .089$ $\eta_p^2 = 0.044$, and an interaction effect, $F(3, 144) = 3.27$, $p = .023$ $\eta_p^2 = 0.064$.

Critically however, the analysis of the *Situational Causation* measure did not reveal a main effect of condition $F < 1$. There was a main effect of scenario, $F(3, 140) = 7.16$ $p = .001$ $\eta_p^2 = 0.131$, suggesting that the situational constraint was higher some scenarios than others, but no interaction, $F < 1$. (Fig. 9)

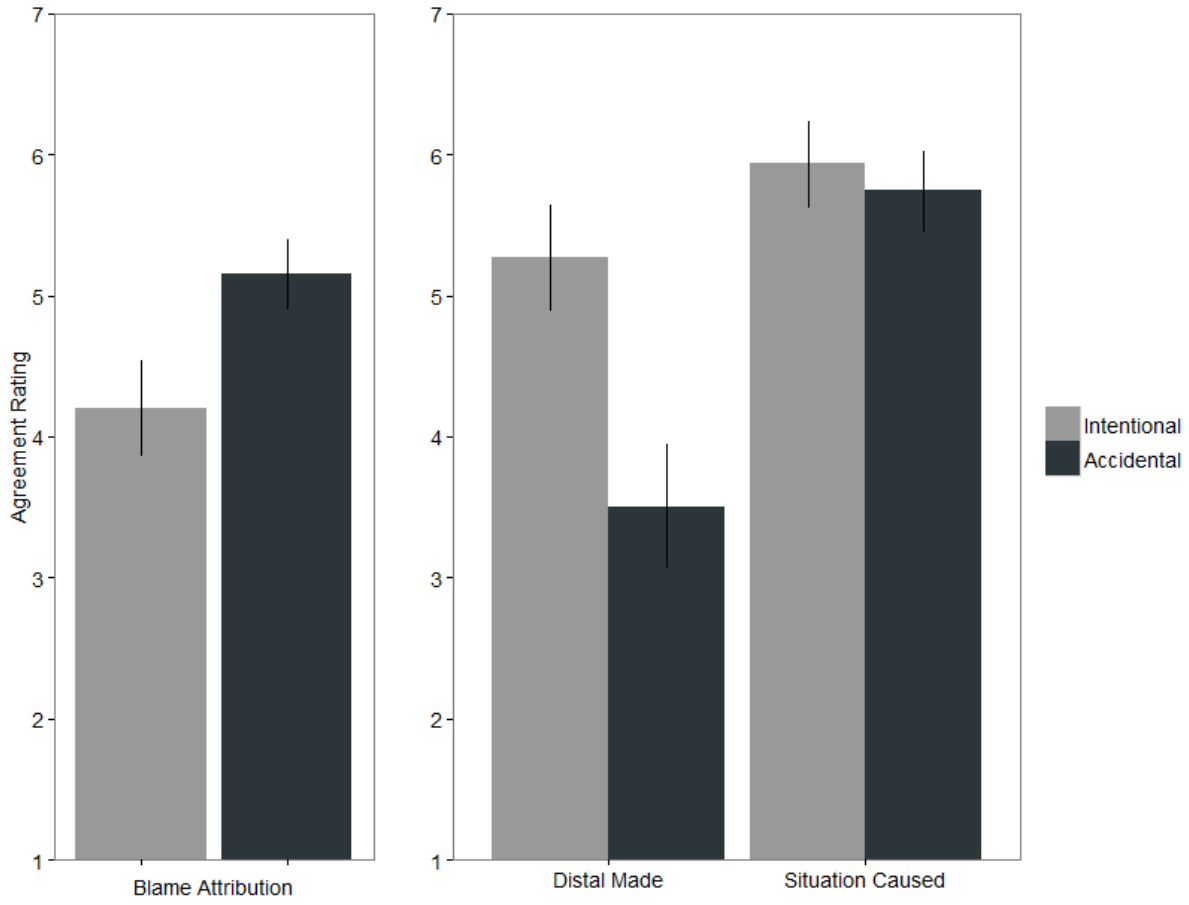

Figure 9. Mean Ratings of Agent Blame, Distal Agent Made and Situation Caused in Intentional vs. Accidental Conditions across All Four Scenarios. Error Bars Indicate Bootstrapped 95% CIs.

As in Study 4a and 4b, the *Distal Agent Causation* measure partially mediated the effect of condition on attributions of blame to the proximal agent (bootstrapped indirect effect = -.399, bias corrected 95% CI: [-.805, -.111]). By contrast, the *Situation Causation* measure did not (bootstrapped indirect effect = -.035, bias corrected 95% CI: [-.196, .025].

7.3 Discussion

These results help clarify the relationship between the third-party's intentions and changes in participants' causal reasoning. Participants' judgments of the causal influence of the *situation* did not differ based on whether or not the distal agent created the situational constraint intentionally.

In addition, this measure of how effective the distal agent was in creating the situational constraint did mediate the reduction in blame to the proximal agent. As such, the perception that the distal agent controlled the proximal agent cannot be accounted for by changes in the situational constraint faced by the proximal agent. On the other hand, participants' causal judgments of the *distal agent* did differ based on the distal agent's intentions, and moreover, these causal judgments mediated the reduction in blame for the proximal agent. Putting these two pieces together: Participants' reduction in blame for the proximal agent seems to be best explained by their perception that the distal agent was controlling the proximal agent, even though this took place without direct interaction between the agents, happened from a distance, and did not involve any change in the situation faced by the proximal agent.

8. General discussion

Taken together, these five studies demonstrate that people's moral responsibility judgments are sensitive to the intentions of third parties and provides evidence that third-party intentions influence judgments of moral responsibility by changing participants' causal reasoning. Study 1 demonstrated that third-party intentions influence judgments of moral responsibility even when the agent lacks all knowledge of third party's intentions. Study 2 found that third-party intentions only influenced moral judgment when there was a precise match between the proximal agent's actions and the distal agent's intention. In Study 3, we provided evidence that third-party intentions exerted their influence on moral responsibility judgments by changing participants' causal reasoning about the agents involved. Study 4 then compared this causal reasoning account to one according to which the effect arose from changes in how blame was distributed between the agents, and provided further evidence for the causal account. Finally, Study 5 compared two

causal mechanisms and further characterized the change in participants' causal reasoning. When the proximal agent's actions matched the distal agent's intentions, the proximal agent was perceived as controlled from a distance by the distal agent.

The finding that third-party intentions impact moral judgments is straightforwardly relevant to the growing body of research on the role of intentions in moral responsibility. This research has consistently focused solely on the intentions of agent who does the immoral act (e.g., Cushman, 2008; Cushman, Young & Hauser 2006; Guglielmo & Malle, 2010a; 2010b; Knobe 2003; Young et al., 2010). In contrast, the present study provides evidence that intentions actually occupy a more pervasive role in moral cognition: they not only influence how people judge individual moral agents, they can also alter people's causal reasoning. That is, even when the situational constraint an agent faces is the same, people may interpret that agent as more 'controlled' (less having the ability to *not* do the immoral action) when the distal agent intentionally, as opposed to accidentally, imposes this constraint. This finding is especially relevant given that the majority of morally valenced actions do not occur as a completely isolated dyad of perpetrator and victim but rather within a complex social and causal network involving many different agents (DeScioli & Kurzban, 2013).

Moreover, the finding that third-party intentions influence moral responsibility by changing causal reasoning relates to two separate bodies of research. First, intentions have played a key role in recent research on causal reasoning (Lombrozo, 2010). Secondly, the connection between causation and moral responsibility has been investigated in recent studies involving causal chains (e.g., Lagnado & Channon, 2008) and has had an important place in the discussion of manipulation and moral responsibility (e.g., Mele, 2006; Nahmias, Coates & Kvaran, 2007; Pereboom, 2001; Rosen 2002).

8.1 Causation and moral responsibility

Previous research examining third-party intentions and causal reasoning has focused on causal chains in which multiple agents independently contributed to the occurrence of some (typically negative) outcome. Using these sorts of causal chains, researchers have examined whether the intentions of the agent who acted first (a distal agent) impacted causal judgments of the more proximal agent who independently causally contributed to the outcome. Interestingly, this research found that causal judgments were not influenced by whether or not the distal agent intended to bring about the eventual outcome (Brickman, Ryan & Wortman, 1975; Hilton et al., 2010; McClure, Hilton & Sutton, 2007).

In similar cases, researchers also failed to observe effects of distal agent intentions on moral judgments of proximal agents. Lagnado and Channon (2008) presented participants with cases in which two independent agents contributed to a negative outcome and asked them to make *both* casual and moral judgments. For example, two agents causally contributed to the death of a sick man: First, his wife (the distal agent) intentionally gave him an overdose of medication and then called the ambulance. Second, the ambulance center (the proximal agent) intentionally ignored the call. As a result, the man died. Employing cases like this, Lagnado and Channon (2008) systematically varied whether the distal and the proximal agent acted intentionally (e.g., the man's wife could have accidentally given him an overdose). Interestingly, while they did observe an effect of the agent's location in the causal chain (the proximal agent was always blamed slightly more than the distal agent), they did not observe an interaction effect between location and intention, suggesting that the proximal agent was not blamed less when the distal agent acted intentionally.

At first blush, the present results may seem to present conflicting evidence. However, the differences in the scenarios employed by Lagnado and Channon may help to explain why we observed an effect of the distal agent's intentions on judgments of moral responsibility for the proximal agent. Specifically, in Lagnado and Channon's study, the distal agent did not act with the intention of bringing about the proximal agent's action, but only with the intention of bringing about the eventual outcome (e.g., the man's wife intended for the man to die, but did not intend for the ambulance center to ignore the call). By contrast, the present research involves cases in which the distal agent's intentions are specifically focused on having the proximal agent do some particular action.

For similar reasons, the present research also differs from previous work using economic games (Bartling & Fischbacher, 2012; Coffman, 2011 Hamman et al., 2010) which indirectly manipulated information about the third party's intentions but did not investigate cases in which the third party specifically tried to control the intermediary agent's actions by implementing a situational constraint. Accordingly, the effect discussed here suggests an additional direction for future research.

8.2 Teleological causal reasoning

Our basic proposal has been that third-party intentions can alter moral responsibility judgments by changing causal reasoning. This proposal builds on previous research in both moral and non-moral causal chains which suggested that the distal agent's intentions influence whether the distal agent is perceived as causing the outcome (Fincham & Shultz, 1981; Lombrozo, 2010; Lombrozo & Carey, 2006). For example, using cases of double prevention (in which an agent prevents an outcome from being prevented from occurring) Lombrozo (2010) found that participants were more inclined to perceive the distal agent as causing the outcome when the

distal agent intentionally (vs. accidentally) acted so as to prevented the outcome from being prevented. Lombrozo explained this effect by arguing that when the outcome matched the intention of the agent who initiated the series of events, participants reasoned about the events *teleologically*. That is, they were approaching the events by thinking about their purpose – the end for which they exist – or what Aristotle called their *final cause* (Barnes, 1984). In contrast, when the outcome did not match the intention of the agent who initiated the series of events, participants were instead reasoning about the events *mechanistically*, focusing on the simple cause and effect relations between any two events, often called 'billiard ball' causation (Hume, 1748/1999).

While the present research differed by focusing on whether the distal agent caused the proximal agent's action (not the eventual outcome of the series of events), future research should investigate whether Lombrozo's proposed explanation can be extended to account for the effect demonstrated in the present research.

8.3 Real world applications

The relevance of third-party intentions is not limited to hypothetical or philosophical cases of manipulation. One commonplace example of the importance of third-party intentions can be found in cases of entrapment (Carlon, 2007), first introduced during prohibition in the case *Sorrells v. United States* (1932). A defense which appeals to entrapment must successfully argue that the government agent induced someone to commit a crime that he or she otherwise would not have committed. Accordingly, the legal discussion of entrapment has focused on the original intentions of the proximal agent who commits the immoral act (*Sorrells v. United States*, 1932; *Sherman v. United States*, 1958) and whether or not the situational constraint created was so

strong that a person not disposed toward crime would be enticed by the situation (Lombardo, 1995). In contrast to this focus, the present research suggests that people's moral judgments are not based solely on the intentions of the agent who committed the crime, but also based on whether the government agent created the situational constraint with the intention of causing the crime to be committed. Jurors may hold the defendant to be more blameworthy if the situational constraint that led to the crime were inadvertently created by law enforcement agents who did not have the intention of causing a crime to be committed.

9. Conclusions

The present study investigated the mechanism underlying the effect of third-party intentions on judgments of moral responsibility and found evidence that a distal agent's intentions affect moral responsibility judgments for proximal agents by altering participants' causal reasoning. Continued research on the role of third-party intentions will not only allow for a better understanding of moral responsibility as an interpersonal phenomenon, but will provide further insight into the relationship between intention, causation and moral responsibility.

Acknowledgements

Among many others, we would like to gratefully acknowledge Paul Bloom, Fiery Cushman, Meghan Freeman, Tania Lombrozo, Dylan Murray, Eddy Nahmias, Shaun Nichols, Liane Young and especially Joshua Knobe for helpful discussion and comments.

References

Alicke, M. (2008). Blaming badly. *Journal of Cognition and Culture*, 8, 179-186. doi:10.1163/156770908X289279

Aristotle. (2002). *Nicomachean ethics*. (S. Broadie & C. Rowe, Trans.). Oxford: Oxford University Press.

Barnes, J. (Ed.) (1984). *The complete works of Aristotle. The revised Oxford translation*. Princeton: Princeton University Press.

Bartling, B. & Fischbacher, U. (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies*, 79, 67-87. doi:10.2139/ssrn.1166544

Borg, J.S., Hynes, C., Van Horn, J., Grafton, S. & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgment: an fMRI investigation. *Journal of Cognitive Neuroscience*, 18, 803–17. doi:10.1162/jocn.2006.18.5.803

Brickman, P., Ryan, K., & Wortman, C. B. (1975). Causal chains: Attribution of responsibility as a function of immediate and prior causes. *Journal of Personality Social Psychology*, 32, 1060-1067.

Buhrmester, M. D., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of cheap, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5. doi:10.1177/1745691610393980

Carlon, A. (2007). Entrapment, Punishment, and the Sadistic State. *Virginia Law Review*, 93, 1081-1134.

Coffman, L. (2011). Intermediation reduces punishment (and reward). *American Economic Journal: Microeconomics*, 3: 77-106. DOI: 10.1257/mic.3.4.77

Cushman, F.A. (2008). Crime and Punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108, 353-380. doi:10.1016/j.cognition.2008.03.006

Cushman, F.A. & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, 35, 1052-1075. doi:10.1111/j.1551-6709.2010.01167.x

Cushman, F., Young, L., Hauser, M. (2006). The role of conscious reasoning and intuitions in moral judgment: Testing three principles of harm. *Psychological Science*, 17, 1082-1089. doi:10.1111/j.1467-9280.2006.01834.x

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139, 477-496.

Fincham, F. D., & Shultz, T. R. (1981). Intervening causation and the mitigation of responsibility for harm. *British Journal of Social Psychology*, 20, 113-120.

Fisher, M. (2004). Responsibility and manipulation. *Journal of Ethics*, 8: 145-77.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, 59, 93-104. doi:10.1037/0003-066X.59.2.93

Greene, J. D. & Cohen J. D. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B*, 359, 1775-17785. doi:10.1098/rstb.2004.1546

Guglielmo, S., & Malle, B. F. (2010a). Can unintended side-effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin, 36*, 1635-1647. doi:10.1177/0146167210386733

Guglielmo, S., & Malle, B. F. (2010b). Enough skill to kill: Intentionality judgments and the moral valence of action. *Cognition*, 117, 139-150. doi:10.1016/j.cognition.2010.08.002

Hamilton, V. L. (1978). Who is responsible? Towards a social psychology of responsibility attribution. *Social Psychology, 41*, 316–328.

Hamman, J. R., Loewenstein, G., & Weber, R. A. (2010). Self-interest through delegation: An additional rationale for the principal-agent relationship. *The American Economic Review*, 100, 1826-1846.

Hart, H. L. A., Honoré, T. (1959/1985). *Causation in the law*. Oxford: Clarendon Press.

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, 383–400. doi:10.1002/ejsp.623

Hume, D. (1998). *Enquiry concerning the principles of morals*. Oxford: Clarendon. (Original work published in 1751)

Hume, D. (1999). *An enquiry concerning human understanding*. Oxford: Oxford University Press. (Original work published in 1748)

Kenny, A. (1973). The history of intention in ethics. In A. Kenny (Ed.), *The anatomy of the soul: Historical essays in the philosophy of mind* (pp. 129–146). Oxford: Blackwell.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63, 190-193.

Kutner, M.H., Nachtsheim, C.J. & Neter, J. (2004). *Applied Linear Regression Models, 4th edition*, McGraw-Hill Irwin.

Leslie, A., Knobe, J. & Cohen, A. (2006). Acting intentionally and the side-effect effect: 'Theory of mind' and moral judgment. *Psychological Science*, 17, 421-427. doi:10.1111/j.1467-9280.2006.01722.x

Lagnado, D. A. & Channon, S. (2008). Judgments of cause and blame: The influence of intentionality and foreseeability. *Cognition*, 108, 754-770. doi:10.1016/j.cognition.2008.06.009

Lombardo, J. (1995). Causation and "objective" entrapment: Toward a culpability-centered approach. *UCLA Law Review*, 43, 209-261.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, 61, 303-332. doi:10.1016/j.cogpsych.2010.05.002

Lombrozo, T. & Carey, S. (2006). Functional explanation and the explanation of function. *Cognition*, 99, 167-204. doi:10.1016/j.cognition.2004.12.009

McClure, J. L., Hilton, D. J., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European Journal of Social Psychology*, 37, 879–901. doi:10.1002/ejsp.394

McKenna, M. (2008). A hard-line reply to Pereboom's four-case manipulation argument. *Philosophy and Phenomenological Research,* 77: 142–159. doi:10.1111/j.1933-1592.2008.00179.x

Mele, A. (1995). *Autonomous agents*. New York: Oxford University Press.

Mele, A. (2006). *Free will and luck*. New York: Oxford University Press.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. New York: Cambridge University Press.

Nadelhoffer, T. (2006). On trying to save the simple view. *Mind & Language,* 21, 565-586. doi:10.1111/j.1468-0017.2006.00292.x

Nahmias, E., Coates, J. & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy,* 31, 214–242. doi:10.1111/j.1475-4975.2007.00158.x

Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes,* 109, 134-141.

Pereboom, D. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.

Pereboom, D. (2008). A hard-line reply to the multiple-case manipulation argument. *Philosophy and Phenomenological Research,* 77: 160-170.

Phelan, M. & Sarkissian, H. (2008). The folk strike back: Or, why you didn't do it intentionally, though it was bad and you knew it. *Philosophical Studies,* 138, 291-298. doi:10.1007/s11098-006-9047-y

Pizarro, D., Uhlmann, E. & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology, 39*, 653-660.

Piaget, J. (1965). *The moral judgment of the child*. New York: Free Press.

Pizarro, D.A. & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (99-108). Washington, DC: American Psychological Association.

Pizarro, D. A., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science*, 14, 267-272.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879-891. doi:10.3758/BRM.40.3.879

Rosen, G. (2002). The case for incompatibilism. *Philosophy and Phenomenological Research*, 64, 699–706.

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

Sherman v. United States, 356 U.S. 369, 372, (78 S.Ct. 819 1958).

Sorrells v. United States, 287 U.S. 435, 451, 53 (S.Ct. 210, 212 1932).

Sripada, C. S. (2012). What makes a manipulated agent unfree? *Philosophy and Phenomenological Research*. 85, 563-593. doi:10.1111/j.1933-1592.2011.00527.x

Uttich, K. & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116, 87-100. doi:10.1016/j.cognition.2010.04.003

Waldmann, M. & Dieterich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. Psychological Science, 18 (3), 247-253.

Waldmann, M., Nagel, J. & Weigmann, A. (2012). Moral judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 364-389). New York, NY: Oxford University Press.

Wiegmann, A., & Waldmann, M. (2014). Transfer effects between moral dilemmas: A causal model theory. *Cognition*. Advanced online publication.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford.

Williams, B. (1993). Recognizing responsibility. In B. Williams (Ed.), *Shame and necessity* (pp. 50–74). Berkeley, CA: University of California Press.

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, 283-301. doi:10.1016/j.cognition.2005.05.002

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron, 65,* 845-851. doi:10.1016/j.neuron.2010.03.003

Young, L. & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuro Image*, 40, 1912-1920. doi:10.1016/j.neuroimage.2008.01.057

Young, L., Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition,* 120, 202-214. doi:10.1016/j.cognition.2011.04.005

Young, L., Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass*. 7: 585–604. doi:10.1111/spc3.12044