Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities

Jonathan Phillips*

Jamie B. Luguri*

Joshua Knobe

Yale University

* These authors contributed equally to this manuscript

Please address editorial correspondence to:

Jonathan Phillips
Department of Psychology
Harvard University
33 Kirkland Street
Cambridge, MA 02138

Email: phillips01@g.harvard.edu

Word count: 10,822

Key words: *morality, alternative possibilities, modality*

Abstract

Past work has demonstrated that people's moral judgments can influence their judgments in a number of domains that might seem to involve straightforward matters of fact, including judgments about freedom, causation, the doing/allowing distinction, and intentional action. The present studies explore whether the effect of morality in these four domains can be explained by changes in the relevance of alternative possibilities. More precisely, we propose that moral judgment influences the degree to which people regard certain alternative possibilities as relevant, which in turn impacts intuitions about freedom, causation, doing/allowing, and intentional action. Employing the stimuli used in previous research, Studies 1a, 2a, 3a, and 4a show that the relevance of alternatives is influenced by moral judgments and mediates the impact of morality on non-moral judgments. Studies 1b, 2b, 3b, and 4b then provide direct empirical evidence for the link between the relevance of alternatives and judgments in these four domains by manipulating (rather than measuring) the relevance of alternative possibilities. Lastly, Study 5 demonstrates that the critical mechanism is not whether alternative possibilities are considered, but whether they are regarded as *relevant*. These studies support a unified framework for understanding the impact of morality across these very different kinds of judgments.

Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities

A series of recent studies have shown that people's moral judgments can impact their intuitions about issues that might appear to be straightforward matters of fact. This effect was noted early on for intuitions about whether an agent acted *intentionally* (Knobe, 2003), but it soon became clear that a very similar pattern could be found in numerous other domains. Among other things, people's moral judgments also influence their intuitions about whether an agent acted *freely*, whether an action *caused* some further outcome, and whether an agent *did* something or merely *allowed* it to happen (Cushman, Knobe & Sinnott-Armstrong, 2008; Hitchcock & Knobe, 2009; Phillips & Knobe 2009; Young & Phillips, 2011).

One way to explain these phenomena would be to offer an independent account for each of the separate effects. So one could construct a hypothesis about why moral considerations influence intuitions about intentional action (Guglielmo & Malle, 2010; Machery, 2008; Nichols & Ulatowski, 2007; Uttich & Lombrozo, 2010), then another separate hypothesis about why moral considerations impact intuitions about causation (Menzies, 2010; Sytsma, Livengood, & Rose, 2012), and so on. (Some have even argued explicitly that there cannot be a unified explanation of all of these effects; Hindriks, 2014.)

Here, we pursue the exact opposite approach. Rather than looking separately at each individual effect, we want to consider the widespread influence of moral cognition in these different domains, and to offer a unified explanation that applies to them all. Specifically, we will suggest that all these effects can be explained in terms of a very general principle involving the relationship between people's moral judgments and their way of understanding *alternative possibilities*.

The Role of Possibilities

Research in a number of different fields has independently argued for the importance of

alternative possibilities, including work in philosophy (e.g., Bennett, 2003; Lewis, 1973), linguistics (Kratzer, 2012; for a review, see Portner, 2009), psychology (Kahneman & Miller, 1986; Byrne, 2005, for a review see, Roese, 1997), and computer science (e.g., Bello, in press; Halpern & Pearl, 2005; Pearl, 2000). In each of these cases, the central insight has been that our understanding of the things that occur is shaped in some fundamental way by our understanding of alternative possibilities that could have occurred but actually did not.

Intriguingly, existing theoretical work has already suggested that alternative possibilities play a role in each of the domains in which we find these surprising effects of moral judgment. As one example, consider judgments about *freedom*. Within existing theoretical work, it is widely acknowledged that the question as to whether someone acted freely depends, in part, on whether it would have been possible for that person to have done something else instead (Aquinas, 1273/1920; Berlin, 1969/2002; Aristotle, 350 BC/2002). Thus, on one view, claims about freedom like (1a) depend on claims about alternative possibilities like (1b).

(1) a. She did *x* freely.

b. If she had wanted not to do *x*, she would not have done *x*.

Similarly, within existing work on judgments about *causation*, it is widely believed that causal judgments in some way rely on representations of alternative possibilities (Lewis, 1973; Pearl, 2000). In particular, a number of accounts of causation suggest that causal claims like (2a) depend at least partially on claims about alternative possibilities like (2b).

(2) a. Event x caused event y.

b. If x had not happened, y would not have happened.

Within existing work in this tradition, there is a great deal of controversy about precisely how to spell out the relationship between alternative possibilities and judgments about freedom, causation, etc. (Halpern & Hitchcock, 2014; Nduibuisi & Byrne, 2013; Woodward, 2004). However, the details

of this controversy will not concern us here. The key point is simply that judgments in each of these domains depend in some way on how we understand the alternative possibilities that could have happened, but didn't.

Across a number of fields, existing work on alternative possibilities has also introduced a further idea that plays a central role in the present hypothesis: People do not treat all alternative possibilities equally. Instead, they regard certain possibilities as *relevant*, while treating others as completely *irrelevant* (Halpern & Hitchcock, 2014; Lewis, 1973; Portner, 2009; Roese, 1997). To the extent that they regard an alternative possibility as irrelevant, they will tend to ignore it entirely, and it will play little role in their judgments in any domain.

To give one simple example, suppose that a group of people is giving a presentation, and they end up doing a terrible job. As we consider what actually occurred in this case, we might regard certain alternative possibilities as especially relevant. For example, possibilities in which they had spent more time preparing might seem especially relevant. Or perhaps even possibilities in which they had simply decided not to give the presentation at all. But then there are numerous other possibilities that we would regard as completely irrelevant. Thus, we would never regard as relevant the possibility that the presentation could have been interrupted by a freak tornado, or the possibility that the earth's gravitational field could have suddenly ceased to exist. People might be capable of entertaining possibilities like these if forced to, but all the same, they would regard them utterly pointless, not even worthy of the slightest consideration.

Most importantly for present purposes, the distinction between relevant and irrelevant possibilities has been invoked to explain judgments in the specific domains under consideration here. Take the example of causation. As we noted above, it seems that people only regard a factor as causal if, when considering alternative possibilities in which that factor does not occur, the outcome also does not occur (Lewis, 1973; Pearl, 2000). But theoretical work suggests that it is not enough

for this to be true; possibilities in which this factor does not occur must also be *relevant*. In cases where such possibilities are regarded as irrelevant, people will simply reject the corresponding causal claim (e.g., Halpern & Hitchcock, 2014; Schaffer & Blanchard, in press). To illustrate this point, consider the causal claim (3a) and the corresponding claim about an alternative possibility (3b).

(3) a. The earth's gravitational field caused that disastrous presentation.

b. If the earth's gravitational field had not been present, the disastrous presentation would not have occurred.

The claim (3b) is surely true. Yet at the same time, possibilities in which the earth's gravitational field is not present seem entirely irrelevant, and thus the theory predicts that the corresponding causal claim should also be seen as absurd.

As this example helps to make clear, it is critical to distinguish between the degree to which people actively *consider* an alternative possibility and the degree to which they regard that alternative possibility as *relevant*. Previous work in social psychology has explored the impact of counterfactual reasoning, and this work has demonstrated that actively considering a particular possibility can have an important impact on many aspects of human behavior (for reviews, see Byrne, 2005; Epstude & Roese, 2008; Roese, 1997). The focus of the present paper, however is on a somewhat different phenomenon. Independent of the question of whether or not people consider a particular possibility, there seems to be a question as to whether people regard that possibility as relevant. (Even if people are specifically instructed to reason about the possibility that the earth's gravitational field could suddenly disappear, they might continue to regard this possibility as completely irrelevant.) As far as we know, this approach has not yet been directly tested in experimental work, but it has been developed in considerable technical detail within the existing theoretical literature in several different fields (e.g., Bello, in press; Halpern & Hitchcock, 2014; Knobe, 2010; Kratzer, 2012; Schaffer & Blanchard, in press).

If we are seeking a unified way of explaining why people's moral judgments influence their judgments in the domains of freedom, cause, etc., then the relevance of alternative possibilities may help. That is, if we can show that people's moral judgments influence their judgments of the relevance of alternative possibilities, then we may be able to provide a unified account of morality's influence in all four of these areas. We propose to test this basic model (Fig. 1) in the current paper.

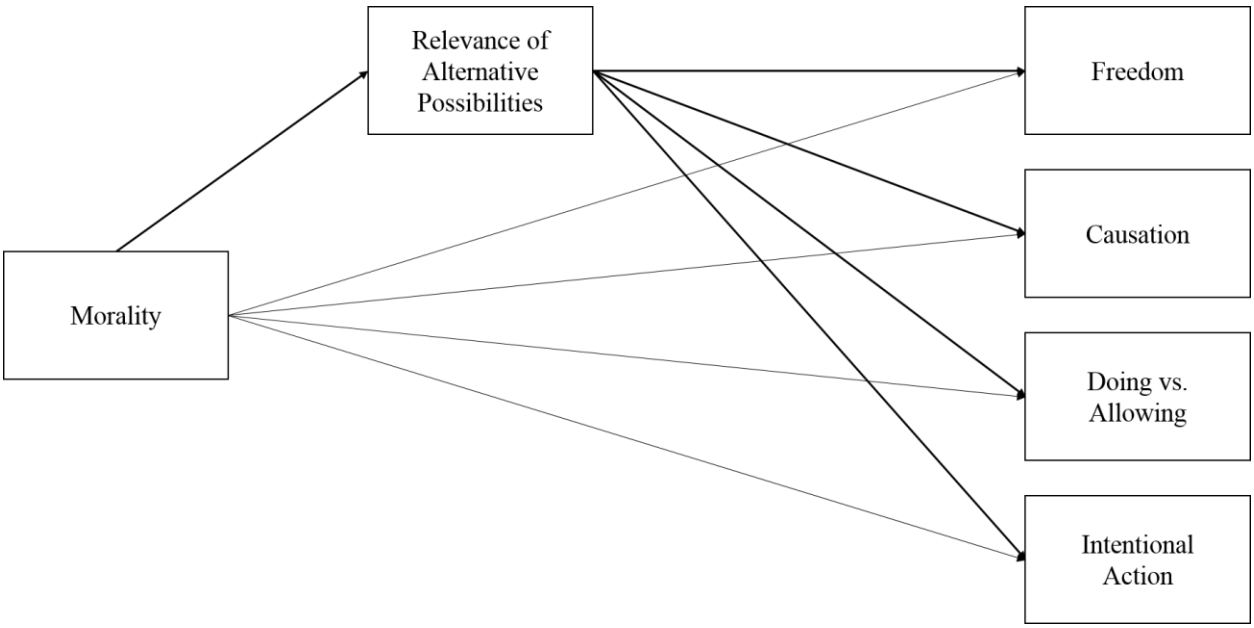

Figure 1. Proposed model with the relevance of alternative possibilities mediating the influence of morality in the domains of freedom, causation, doing vs. allowing, and intentional action.

Morality and the Relevance of Alternative Possibilities

Thus far, we have been considering the evidence that judgments of the relevance of alternative possibilities influence intuitions in each of the four domains. We now turn to the other link in our model – the claim that moral judgments influence judgments about the relevance of alternative possibilities. Specifically, there is a general tendency to regard alternative possibilities as more relevant to the extent that they involve replacing something morally bad (in the actual world)

with something morally good (in the alternative possibilities). Some form of this basic idea can be found across a wide variety of different theoretical accounts that have been spelled out within quite different theoretical frameworks (e.g., Halpern & Hitchcock, 2014; Knobe & Szabo, 2013; Schaffer & Blanchard, in press). The precise technical implementation varies considerably from one framework to the next, but the core idea that is shared across all of these accounts is a highly intuitive one.

Suppose you believe that a certain aspect of the way people typically treat each other is fundamentally morally wrong. You might then see it as highly relevant to consider alternative possibilities in which people treated each other in the way you believed to be morally good. If you saw a man insulting a homeless person, for example, the alternative that he could have instead tried to help the homeless person clearly seems relevant. Now, by contrast, suppose you believe that a certain aspect of the way people typically treat each other is morally good. In that latter case, you might regard it as completely irrelevant to consider alternative possibilities in which, for some reason, people treated each other in the way you believe to be morally bad. It clearly does not seem relevant to think about how the man could have insulted the homeless man instead of helping him.

The Present Studies

Past work offers theoretical support for the idea that the influence of morality may occur by altering people's intuitions about the relevance of possibilities. Yet to date, there has been little empirical work that directly investigates this relationship. This is our aim in the present studies. More precisely, we propose that morality influences the degree to which people judge certain alternative possibilities as relevant, which in turn impacts people's intuitions in non-moral domains.

To test this model, we took the materials from four previously published studies that originally demonstrated morality's influence in the domains of freedom, causation, doing/allowing, and intentional action. The use of these previous materials naturally limits researcher degrees of

freedom and experimenter bias (Simmons, Nelson, & Simonsohn, 2011; Strickland & Suben, 2012). Further constraining researcher degrees of freedom, we also explored each domain using exactly the same methods. Specifically, within each domain, we conducted two studies: one that checked for mediation by relevance of alternatives and one that manipulated relevance of alternatives.

For each of the mediation studies, we replicated an existing experiment and then added a further measure in which participants were directly asked about the relevance of certain alternatives. We predicted that the different conditions of the original experiment would lead to different judgments on the measure of relevance of alternatives, which would in turn mediate the impact of condition on the original dependent variable.

For the manipulation studies, we took the morally neutral condition of each of the original experiments and then tried to manipulate the relevance of alternatives using an entirely non-moral method. If morality affects judgments of freedom, causation, doing/allowing, and intentional action by changing the relevance of alternative possibilities, then morality should not be special in its influence on these domains. In fact, any factor that influences the relevance of alternative possibilities should also impact judgments in these various domains. Accordingly, to manipulate the extent to which participants regarded alternatives to the agents' actions as relevant, participants were instructed to write a brief paragraph about what else the agent could have done. (Participants in a control condition were simply asked to summarize the story). Admittedly, the effect of this novel way of manipulating the relevance of alternative possibilities may not be as strong as the original effects of morality observed in the previous studies. Still, as this manipulation parallels the proposed mechanism, we expect that participants who are asked to engage in writing about alternative possibilities (vs. summarize the story) will show a pattern of judgments similar to those observed in the morally bad (vs. morally neutral) actions.

These experimental methods are applied to judgments of freedom (Study 1a-b), causation

(Study 2a-b), doing vs. allowing (Study 3a-b), and intentional action (Study 4a-b). In each case, we find two pieces of support for the proposed model. First, we find that judgments of the relevance of alternative possibilities mediate the effect of morality on judgments in these four domains. However, the evidence from these mediation analyses is equivocal because the data fit alternative mediation models. Second, and more conclusively, we provide direct evidence of the role of alternative possibilities by non-morally manipulating the relevance of alternatives and then demonstrating that all four kinds of judgments are affected by this manipulation in precisely the same way that they were affected by morality. Having considered relevant alternatives to the agent's action, participants judge that the actor was less forced, more of a cause, did something rather than merely allowed it to happen, and acted more intentionally. Study 5 then provides a demonstration that it is the perceived *relevance* of alternatives (rather than simply the process of considering of alternative possibilities) that affects participants' judgments. Taken together, these five studies provide support for a unifying framework in which the impact of morality on a wide array of disparate judgments can be accounted for by morality's influence on the relevance of alternative possibilities.

Study 1: Freedom and Force

Past research has found that morality impacts judgments about whether an agent performed an action *freely* or was simply *forced* to perform that action (Phillips & Knobe, 2009; Young & Phillips, 2011). In one study, participants were asked to read a scenario in which a ship captain saved his ship from sinking either by doing something morally bad (throwing his wife overboard) or by doing something morally neutral (throwing his wife's cargo overboard). Participants were less likely to judge that the ship captain was forced to throw something overboard when doing so was morally bad than when it was not.

In theoretical accounts of freedom (Aquinas, 1273/1920; Berlin, 1969/2002; Aristotle, 350

BC/2002), whether or not an agent is rightly said to have acted freely depends critically on what *other* actions the agent could have done instead. It has proven remarkably difficult to say precisely how it is that the concept of freedom relates to these other possible actions, and which other possibilities bear on the question as to whether the agent acted freely (see, e.g., Cova, 2014; Cova & Kitano, 2014; Miller & Feltz, 2011; Woolfolk, Doris & Darley, 2006). However, what we rely on here is simply the idea that alternative possibilities play some essential role in people's judgments about freedom.

Thus, one possible way to explain the impact of moral judgments on judgments of freedom is to argue that participants' moral judgments influenced the extent to which they found it relevant to consider the alternative actions that the agent could have done instead. We test this possibility using both a mediational analysis (Study 1a) and a direct manipulation of the relevance of alternative possibilities (Study 1b).

Study 1a Mediation

If people's moral judgments are changing which alternative possibilities they consider to be relevant when reading the vignette, we predict that participants will judge that it is more relevant to consider the possibility that the ship captain could have not thrown his wife (vs. his wife's cargo) overboard. Moreover, we predict that the previously observed impact of morality on judgments of freedom will be mediated by these judgments of relevance.

Method

Participants. Four hundred participants (148 female, 1 unreported, mean age = 32.65, *SD*=10.07) were recruited on Amazon's Mechanical Turk in exchange for a small monetary payment.

Procedure. All participants read the vignette about the ship captain (Phillips & Knobe, 2009). Those in the morally neutral condition read about the captain throwing his wife's cargo

overboard, whereas those in the morally bad condition read about the captain throwing his wife overboard:

> "While sailing on the sea, a large storm came upon a captain and his ship. As the waves began to grow larger, the captain realized that his small vessel was too heavy and the ship would flood if he didn't make it lighter.
>
> The only way that the captain could keep the ship from capsizing was to throw his wife [his wife's expensive cargo] overboard. Thinking quickly, the captain took her [her cargo] and tossed it into the sea. While the captain's wife [the expensive cargo] sank to the bottom of the sea, the captain was able to survive the storm and returned home safely."

Secondly, participants completed the relevance of alternatives measure. Because participants are most likely unfamiliar with being asked to report on the relevance of alternative possibilities, we designed a measure that presented this basic concept in an easily understandable conversational context. Participants were asked to imagine that two people, Sam and Alex, had the following discussion concerning the scenario:

> Alex: "I wonder how things could have gone differently."
> Sam: "Well, the captain could have decided not to throw the cargo [his wife] overboard."
> Alex: "Really? Of all the ways things could have gone differently, that doesn't seem like the one that's relevant to consider."

Notice that Alex's final statement is a claim that the possibility raised by Sam is an irrelevant one. Accordingly, participants were simply asked to rate how much they agreed with Alex, on a scale from 1 ("completely disagree") to 7 ("completely agree"). This agreement rating was then reverse coded to give us a direct measure of the relevance (rather than irrelevance) of this alternative possibility.

After completing this measure, participants were asked whether they agreed or disagreed with the statement "The ship captain was forced to throw his wife's expensive cargo [his wife] overboard." on a scale from 1 ("disagree") to 7 ("agree"). Lastly, demographic information was

assessed. All stimuli, data, and analyses are available at: https://github.com/phillipsjs/RoA

Results

First, we tested whether morality had an influence on judgments of force. Replicating previous results, we found that participants more agreed that captain was forced in the morally neutral condition (M=5.56, SD=1.50) than in the morally bad condition (M=2.54, SD=1.52), $t(398)$= -20.00, $p < .001$, d = 2.00.

Next, we investigated whether morality influenced how relevant participants found the proposed alternative possibility (the hypothesized mediating variable). As predicted, people were more likely to judge that it was relevant to consider the possibility that the captain could have not thrown his wife overboard (M=5.72, SD=1.68) than they were to judge that it was relevant to consider the possibility that he could have not thrown his wife's cargo overboard (M=3.43, SD=1.97), $t(388.48)$=12.529, p<.001, d=1.253.

We then used a bootstrap mediational analysis (Preacher & Hayes, 2008) to test whether judgments of the relevance of alternatives mediates the effect of morality on judgments of force (Fig. 2). Using 5,000 resamples, we found that there was a significant indirect effect of morality condition on judgments of force through judgments of relevance (95% CIs [0.40, 0.89]).

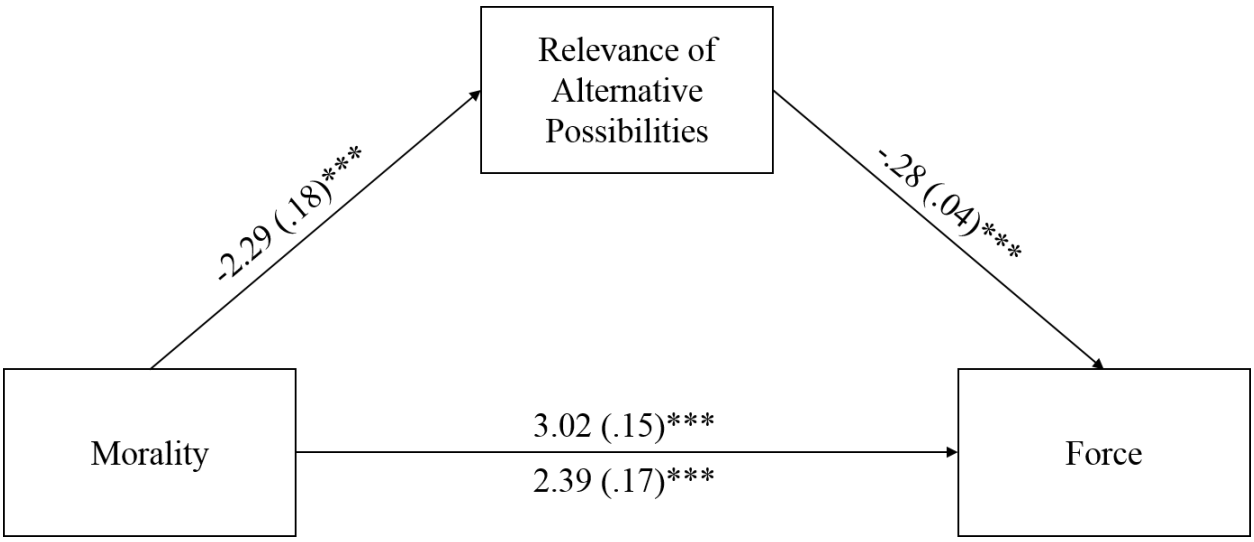

Figure 2. Mediation model with judgments of the relevance of alternative possibilities mediating the effect of morality on judgments of force. Unstandarized coefficients and standard errors (in parentheses) are shown for each path. Asterisks indicate significance, ***p<.001.

We additionally tested the 'reverse' mediation model (with force mediating the impact of condition on relevance) and found that it was also supported by the data, 95% CIs [-1.62, -0.83]. There are at least two possible explanations for the statistical support we find for this alternative model. The first is that the moral valence of the agent's action could have somehow more directly changed participant's judgments of freedom, and then participants' judgments of the relevance of alternative possibilities were affected by this shift in whether the agent was regarded as having acted freely. A second, simpler possibility is that our novel measure of relevance was simply insensitive or noisy. If true, it would be unsurprising that judgments of freedom mediated the effect of morality on the relevance of alternatives, since judgments of freedom themselves have long been understood to be highly sensitive to changes in relevance of alternative possibilities.

Study 1b. Manipulation

To provide a more conclusive test of the role of the relevance of alternative possibilities, we turn to a direct manipulation of this proposed mediator and ask whether this has the same effect as

morality on judgments of freedom. Specifically, we test whether non-morally increasing the relevance of alternative actions the agent could have taken leads participants to judge he acted more freely.

Method

Participants. One hundred and six participants (53 women, mean age=34.14, *SD*=12.14) completed a questionnaire using Amazon's Mechanical Turk.

Procedure. All participants received the neutral case from Study 1a, in which the captain throws his wife's cargo overboard. Participants were assigned either to a control condition or a 'relevant alternatives' condition. Participants in the control condition were simply asked to summarize and describe the events that actually happened in the vignette. Those in the relevant alternatives condition were asked to think about what other decisions the ship captain could have made.

Participants were then asked whether they agreed or disagreed with the claim that the captain was forced to throw the cargo overboard. Participants rated their agreement with this statement on a scale from 1 ('completely disagree') to 7 ('completely agree').

Results

Mirroring the effect of morality in Study 1a, participants in the relevant alternatives condition rated the ship captain as significantly less forced (M=5.46, SD=1.50) than those in the control condition (M=6.17, SD=1.59), $t(104)=2.36$, $p=.02$, $d=.46$.

Study 1a-b Discussion

Employing the stimuli used in previous research, we first tested whether the moral valence of an action influences people's judgments of the relevance of alternative possibilities. As hypothesized, when an action was morally bad (vs. morally neutral), people were more likely to judge as relevant the alternative possibilities in which the agent does not do that action, and these

judgments of relevance also mediated the effect of morality on judgments of freedom. The mediation we observed was only partial, which could have resulted either from (1) the relevance of alternative possibilities accounting for a relatively small portion of the effect or (2) a relatively weak relationship between participants' actual representation of alternative possibilities and the particular measure we employed. However, Study 1b provided further support for the hypothesized role of the relevance of alternatives possibilities in the judgments of freedom. After participants generated alternatives to the agent's action (e.g., "He could've thrown his own stuff overboard."), they judged that the actor was less forced to do the action he did, precisely mirroring the pattern observed when morality was manipulated.

Study 2: Causation

As in the work on freedom and force, previous research on causal cognition has found that the moral status of an action also influences the extent to which people judge that action to be a cause of some outcome (Hitchcock & Knobe, 2009; Alicke, 2000). For example, in one study, participants were asked to read a vignette about a philosophy department in which administrative assistants were allowed to take pens, but professors were not. One day, an administrative assistant took one of the two last pens (a morally neutral action), and so did a professor (a morally bad action). A problem then arose because there were no more pens. When asked who caused the problem, participants overwhelmingly agreed that the professor was the cause of the problem (Knobe & Fraser, 2008).

As with freedom, prior theoretical work in the domain of causal cognition has proposed that causal judgments depend in some fundamental way on alternative possibilities (Lewis, 1979; Pearl, 2000). Within the research on this relationship, there are different views on the topic, with some saying that causal cognition should be understood in terms of alternative possibilities (e.g.,

Gerstenberg, Goodman, Lagnado & Tenenbaum, 2015; Wells & Gavanski, 1989), others emphasizing that alternative possibilities themselves are insufficient and arguing for a role for mechanism information (Copley & Wolff, 2014; Walsh & Sloman, 2011), and still others proposing pluralistic views that involve both alternative possibilities and mechanism information (Lombrozo, 2010). As in the case of freedom, the central idea we rely on here is simply that alternative possibilities play *some* role in causal cognition.

Drawing on these accounts, it may be that part of the reason people's moral judgments impact their causal judgments is that people's moral judgments impact the degree to which they regard certain alternatives as relevant, which in turn impacts their causal cognition. Here we directly test this prediction.

Study 2a: Mediation

Method

Participants. Four hundred participants (154 women, mean age = 29.10, *SD*=8.84) were recruited on Amazon's Mechanical Turk in exchange for payment. One participant did not complete the survey and was excluded from the analyses.

Procedure. Participants were asked to read one of two variants of the scenario about the professor and the pens (Knobe & Fraser, 2008). The original study used a within-subjects design that does not allow for mediational analyses. Here, we instead employ a between-subjects manipulation of the morality of Professor Smith's action that allows us to test the hypothesized mediation as in Study 1.[1] Accordingly, Professor Smith's action was described as morally wrong in one condition:

> The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own. The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly e-mailed them reminders that only administrators are allowed to take the pens.

and was described as morally neutral in the other condition:

> The receptionist in the philosophy department keeps her desk stocked with pens. Both the administrative assistants and the faculty members are allowed to take the pens, and both the administrative assistants and the faculty members typically do take the pens. The receptionist has repeatedly e-mailed them reminders that both administrators and professors are allowed to take the pens.

In both conditions, participants were told that the professor and the administrative assistant took pens:

> On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later, that day, the receptionist needs to take an important message... but she has a problem. There are no pens left on her desk.

Then, as in Study 1, participants read about two people (Sam and Alex) discussing how things could have gone differently. Sam raises the possibility that Professor Smith could have not taken the pen, and Alex responds that this possibility is not one that is relevant to consider. Participants were then asked to rate how much they agreed with Alex, on a scale from 1 ("completely disagree") to 7 ("completely agree"). In addition, participants were asked how much they agreed with the statement "Professor Smith caused the problem" on a scale from 1 ("strongly disagree") to 7 ("strongly agree"). Lastly, demographic information was assessed.

Results

In line with the previous results, we found that participants rated Professor Smith as more of a cause of the problem when he acted immorally (M=4.82, SD=1.63) than when his action was morally neutral (M=2.39, SD=1.48), $t(397)=15.58$, $p<.001$, d=1.56. Next, we investigated whether participants judged the possibility that Professor Smith could have not taken a pen to be more relevant when he acted immorally. We again reverse-coded these judgments, so that they indicated how relevant (rather than irrelevant) participant regarded this alternative possibility to be. As predicted, participants regarded the possibility in which the professor did not take a pen to be more relevant when this action was immoral (M=5.08, SD=1.79) than when this action was morally

neutral (M=3.42, SD=1.84), t(397)=9.14, p<.001, d=.915.

We then used a bootstrap mediational analysis to test whether judgments of the relevance of alternatives mediates the effect of morality on judgments of causation (Fig. 3). Using 5,000 resamples, we found that there was a significant indirect effect of morality condition on judgments of force through judgments of relevance (95% CIs [-0.88,-0.44]).

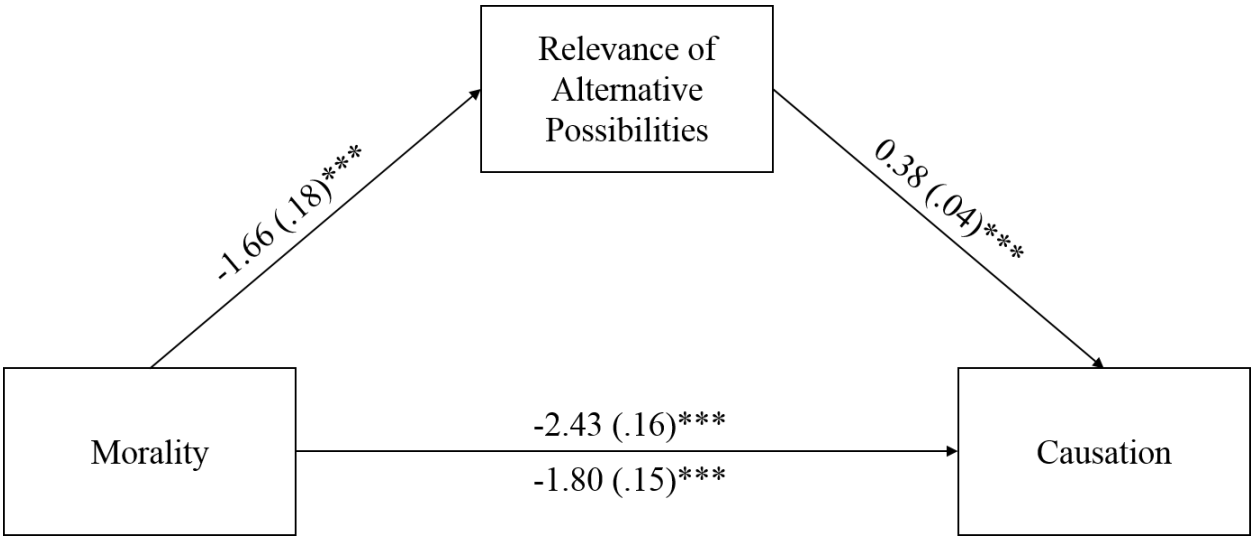

Figure 3. Mediation model with judgments of the relevance of alternative possibilities mediating the effect of morality on judgments of causation. Unstandarized coefficients and standard errors (in parentheses) are shown for each path. Asterisks indicate significance, ***p<.001.

We also tested the 'reverse' mediation model (with causation mediating the impact of condition on relevance) and again found that it was also supported by the data 95% CIs [-1.61, -0.95]. As with judgments of force, the same two explanations may account for the support we find for this alternative mediation model. Accordingly, we now turn to a direct manipulation of the proposed mediator.

Study 2b: Manipulation

Similar to the case of freedom, we test whether non-morally manipulating the relevance of

alternative possibilities impacts participants' judgments in the same way that morality does. More specifically, we ask whether non-morally increasing the relevance of alternative actions the agent could have taken leads participants to see the agent as more of a cause of the outcome.

Method

Participants. Two hundred thirty-four participants (109 women, mean age = 30.98, *SD*=9.24) completed a questionnaire using Amazon's Mechanical Turk.

Procedure. All participants received the morally neutral vignette from Study 2a in which the professor is allowed to take the pens. Participants were assigned either to a control condition, in which they were simply asked to summarize and describe the events that actually happened in the vignette, or a 'relevant alternatives' condition, in which they were asked to think about what other decisions the professor could have made, other than deciding to take a pen.

Participants were then asked whether they agreed or disagreed with the claim that the professor caused the problem. Participants rated their agreement with this statement on a scale from 1 ('completely disagree') to 7 ('completely agree').

Results

Mirroring the effect of morality in Study 2a, participants in the relevant alternatives condition rated the professor as more of a cause (M=3.04, SD=1.80) than those in the control condition (M=2.34, SD=1.45), $t(322.00)=3.31$, p=.001, d=0.42, corrected for unequal variance**.**

Study 2a-b Discussion

As in the domain of freedom, we again found that moral valence of an action influences people's judgments about the relevance of alternative possibilities. Additionally, these judgments mediated the effect of morality on participants' causal judgments. Study 2b then provided further evidence of the role of the relevance of alternatives possibilities: when participants were asked to consider alternatives to the agent's action, they judged the actor to be more of a cause of the

outcome that occurred, mirroring the pattern observed when morality was manipulated in Study 2a.

Study 3: Doing/Allowing

In Study 3, we take up judgments of doing vs. allowing. Previous work has demonstrated that, when an agent performs a morally bad action, people judge that the actor *did* something rather than merely *allowed* it to happen (Barry, Lindauer & Øverland, in press; Cushman, Knobe & Sinnott-Armstrong, 2008). For example, participants were asked to read a vignette about a homeless man who was brought into the hospital because his organ systems were shutting down. He was attached to a ventilator, and would only live for a week or two, regardless of what his doctor did. In the morally neutral condition, the doctor disconnected the man from the ventilator because he wanted to save his dignity. In the morally bad condition, the doctor disconnected the man because he did not want to waste his time and resources on a homeless man. In both cases, the homeless man died as a result of the ventilator being disconnected. Participants are much more likely to say that the doctor ended the man's life (vs. allowed it to end) when they read about the morally bad doctor.

Previous theoretical research has suggested that the distinction between doing and allowing can be spelled out in terms of alternative possibilities (though substantial disagreement remains about how to work out such an account in detail, Bennett 1993; 1995; Kagan, 1989). This emphasis on the role of possibilities in the theoretical research suggests that one promising approach to explaining the impact of morality on judgments doing/allowing may be to consider how morality impacts the relevance of alternative possibilities in such scenarios.

Study 3a: Mediation

Method

Participants. Three hundred and ninety-nine participants (157 women, 2 unreported, mean age=31.62, *SD*=10.50) were recruited on Amazon's Mechanical Turk.

Procedure. Participants were asked to imagine the following scenario (Cushman, et al., 2008):

> Dr. Bennett is an emergency-room physician. An unconscious homeless man is brought in, and his identity is unknown. His organ systems have shut down and a nurse has hooked him up to a respirator. Without the respirator he would die. With the respirator and some attention from Dr. Bennett he would live for a week or two, but he would never regain consciousness and could not live longer than two weeks.

Participants assigned to the morally neutral condition then read:

> Dr. Bennett thinks to himself, "This poor man deserves to die with dignity. He shouldn't spend his last days hooked up to such a horrible machine. The best thing to do would be to disconnect him from the machine."

Whereas participants assigned to the morally bad condition read:

> Dr. Bennett thinks to himself, "This bum deserves to die. He shouldn't sit here soaking up my valuable time and resources. The best thing to do would be to disconnect him from the machine."

Then, participants read that Dr. Bennett proceeded to disconnect the man from the machine, and the man quickly dies. Next, similar to the first two studies, participants were asked to imagine two people (Sam and Alex) discussing how things could have gone differently. Sam raises the alternative possibility in which Dr. Bennett did not decide to disconnect the man from the machine, and Alex objects, indicating that this possibility is not a relevant one. Participants were asked to rate the extent to which they agreed with Alex on a scale from 1 ("completely disagree") to 7 ("completely agree"). As before, this measure was reverse coded to indicate a judgment of relevance, rather than irrelevance. Next, they were asked whether they thought that Dr. Bennett caused the man's life to end, or allowed it to end, on a scale from 1 ("allowed to end") to 7 ("ended"). Lastly, demographic information was assessed.

Results

Replicating previous results, participants in the morally bad condition were more likely to say that Dr. Bennett ended the man's life (vs. allowed it to end; M=4.54, SD=2.29) compared to those in the morally neutral condition (M=3.01, SD=2.02), $t(397)=7.05$, $p<.001$, $d=0.71$. Next, we tested

whether participants rated the possibility that the doctor could have not disconnected the man as more relevant in the morally bad (vs. morally neutral) condition. Indeed, as predicted, we found that participants were more likely to judge the alternative possibility as relevant in the morally bad condition (M=5.30, SD=1.83) than the morally neutral condition (M=4.58, SD=1.95), $t(397)=3.79$, $p<.001$, $d=.38$, suggesting a connection between morality and the relevance of alternative possibilities in a third domain.

We then tested whether these judgments of relevance mediated the relationship between morality condition and judgments of doing/allowing (See Fig. 4). Using bootstrap mediation (5,000 resamples), we found a significant indirect effect through judgments of relevance (95% CIs [-0.35, -0.07]).

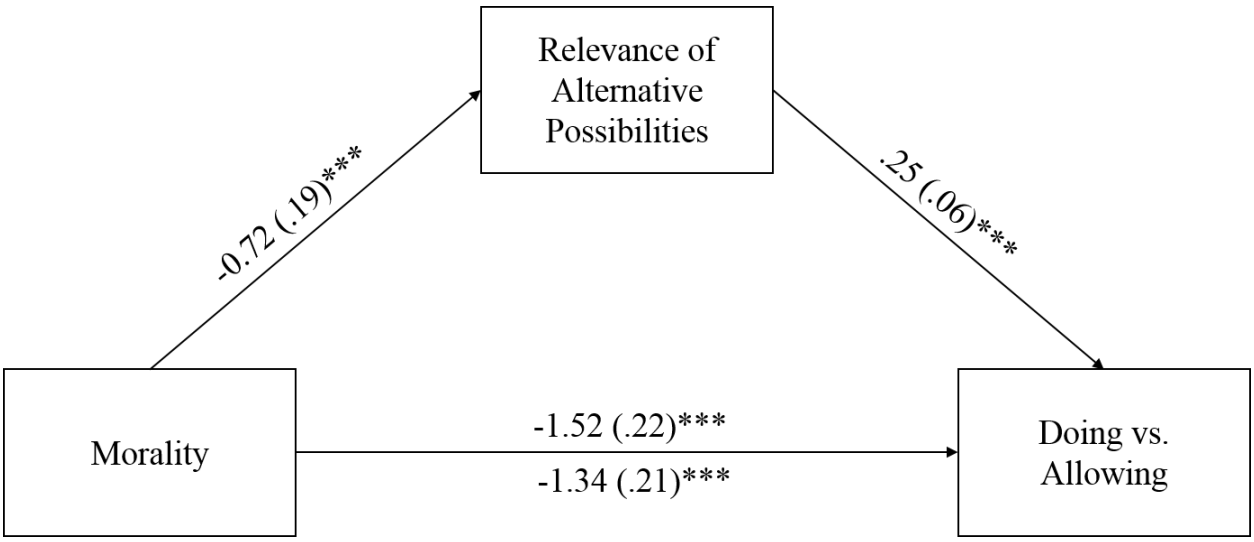

Figure 3. Mediation model with judgments of the relevance of alternative possibilities mediating the effect of morality on judgments of doing vs. allowing. Unstandarized coefficients and standard errors (in parentheses) are shown for each path. Asterisks indicate significance, ***$p<.001$.

We also tested the 'reverse' mediation model and again found that it was also supported by the data 95% CIs [-0.49, -0.16]. Accordingly, we now turn to a direct manipulation of the proposed

mediator.

Study 3b: Manipulation

Similar to the previous studies, we test whether manipulating the relevance of alternative possibilities has the same effect as morality on participants' judgments. Specifically, we ask whether non-morally increasing the relevance of alternative actions the agent could have taken leads participants to view the agent as more having done something rather than allowing it to happen.

Method

Participants. Two hundred thirty-nine participants (86 women, mean age=30.92, *SD*=10.30) completed a questionnaire using Amazon's Mechanical Turk.

Procedure. All participants received the vignette from Study 3a about the doctor who unplugs the ventilator in order to preserve the patient's dignity. Participants were assigned either to a control condition or a 'relevant alternatives' condition. Participants in the control condition were simply asked to summarize and describe the events in the vignette. Those in the relevant alternatives condition were asked to think about what other decisions the doctor could have made, other than deciding to unplug the machine.

Participants were then asked whether they thought that Dr. Bennett caused the man's life to end, or allowed it to end, on a scale from 1 ("allowed to end") to 7 ("ended").

Results

Mirroring the effect of morality in Study 3a, participants in the relevant alternatives condition (M=3.46, SD=2.18) had higher ratings of ending the homeless man's life (vs. allowing it to end), as compared to those in the control condition (M=2.90, SD=1.96), $t(237)$=-2.09, p=.038, d=.27.

Study 3a-b Discussion

As in the domains of freedom and causation, we found the impact of morality on judgments

of doing vs. allowing were mediated by changes in participants' judgments about the relevance of alternative possibilities. Moreover, Study 3b demonstrated that when participants were asked to consider alternatives to the agent's action, they judged the actor to more have done something rather than allowed it to happen, as was the case when morality was manipulated in Study 3a.

Study 4: Intentional Action

We now turn to the effect of moral judgment on intuitions about intentional action. In previous work (Knobe, 2003), participants were asked to read a vignette about the vice president of a company who went to the chairman of the board and proposed a new program that would increase profits, but would also have a side effect. The side-effect was either morally good (helping the environment) or morally bad (harming the environment). In both cases, the chairman said he didn't care about the side-effect – his sole aim is to increase profits. Participants were much more likely to say that the chairman brought about the side-effect intentionally when it was morally bad than when it was morally good.

Existing theoretical work has proposed that this effect, too, can be explained in terms of alternative possibilities (Cova & Egré, in press; Nduibuisi & Byrne, 2013). The key idea is that the agent's actual attitude is exactly the same in both conditions (not caring at all) but this attitude is construed very differently depending on which alternative possibility people consider. In the condition where the agent helps the environment, people tend to consider possibilities in which the agent specifically *wants* to help the environment. Compared to these alternative possibilities, the actual agent seems to show a surprising lack of interest in helping the environment, and his behavior is therefore regarded as unintentional. By contrast, in the condition where the agent harms the environment, people tend to consider possibilities in which the agent specifically *wants to avoid* harming the environment. Compared to these alternative possibilities, the actual agent seems to be

surprisingly willing to harm the environment, and his behavior is therefore regarded as intentional.

In line with this, we test the prediction that when the side-effect is bad, participants will tend to regard as relevant possibilities in which the chairman wanted to avoid the side effect, and that these judgments of relevance explain the impact of morality on intuitions about intentional action.

Study 4a Mediation

Method

Participants. Four hundred and one participants (147 women, mean age = 31.93 *SD*=10.30) were recruited on Amazon's Mechanical Turk.

Procedure. Participants were asked to read the following scenario (from Knobe, 2003):

> The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, and it will also harm [help] the environment."
>
> The chairman of the board answered, "I don't care at all about harming [helping] the environment. I just want to make as much profit as I can. Let's start the new program."
>
> They started the new program. Sure enough, the environment was harmed [helped].

As in the first three studies, participants were asked to imagine two people (Sam and Alex) discussing what happened.

> Alex: "I wonder how things could have gone differently."
> Sam: "Well, the chairman could have wanted to avoid harming the environment [helping the environment]."
> Alex: "Really? Of all the ways things could have gone differently, that doesn't seem like the one that's relevant to consider."

Participants were asked to rate the extent to which they agreed with Alex's judgment that the possibility Sam raised was irrelevant on a scale from 1 ("completely disagree") to 7 ("completely agree"). As in all of the preceding studies, this measure was reverse coded. Next, they were asked whether they agree with the statement that the chairman of the board intentionally harmed the environment/shifted the consumer base on a scale from 1 ("strongly disagree") to 7 ("strongly

agree"). Lastly, demographic information was assessed.

Results

Participants rated the chairman's actions as more intentional when the side effect was harming the environment (M=6.00, SD=1.29), compared to helping the environment (M=1.87, SD=1.28), $t(399) = 32.10$, $p<.001$, $d = 3.21$. Next, we tested whether participants differentially viewed as relevant the possibility that the chairman could have wanted to avoid the side-effect, depending on condition. As predicted, we found that participants were more likely to judge this alternative possibility as relevant when the environment was harmed (M=5.54, SD=1.73), compared to when it was helped (M=3.41, SD=1.80), $t(399)$=12.10, $p<.001$, d=1.21. We then tested whether these judgments of relevance mediated the relationship between morality and judgments of intentional action (See Fig. 5). Using a bootstrap mediation (5,000 resamples), we found a significant indirect effect through judgments of relevance (95% CIs [-0.48, -0.12]).

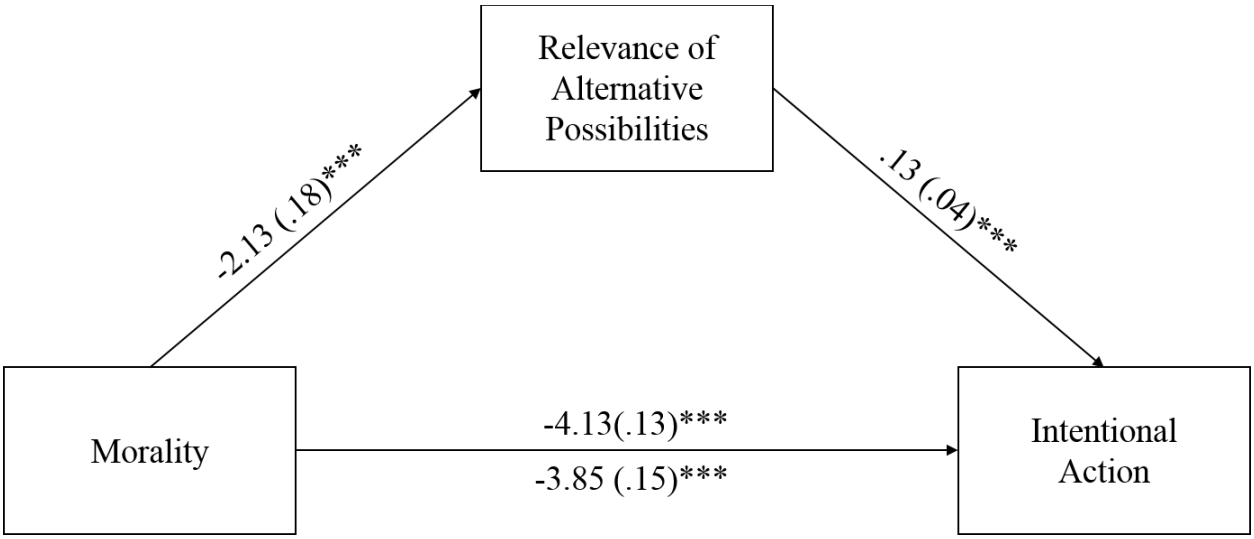

Figure 5. Mediation model with judgments of the relevance of alternative possibilities mediating the effect of morality on judgments of intentional action. Unstandarized coefficients and standard errors (in parentheses) are shown for each path. Asterisks indicate significance, ***$p<.001$.

We also tested the 'reverse' mediation model and again found that it was also supported by the data 95% CIs [-1.58, -0.49]. Accordingly, we now turn to a direct manipulation of the proposed mediator.

Study 4b: Manipulation

Unlike the three other domains, the original research on intentional action did not involve a morally neutral condition, but compared a case in which the side effect was morally bad (harming the environment) to one that is morally good (helping the environment). For consistency with the previous studies, we employ a case in which the side-effect was morally neutral (shifting the consumer base to be slightly older). Drawing on existing theoretical explanations of this effect, we then asked half of the participants to reflect on an alternative possibility in which the chairman was in favor of this neutral side-effect (paralleling the original 'help' vignette), and asked the other half of participants to reflect on an alternative possibility in which the chairman was against shifting the consumer base (paralleling the original 'harm' vignette). As in the previous studies, we test whether manipulating the relevance of alternative possibilities has the same effect as morality on participants' judgments. Specifically, we ask whether non-morally increasing the relevance of the alternative possibility in which the chairman was against shifting the consumer base leads participants to view the chair as having acted more intentionally.

Method

Participants. Two hundred twenty-eight participants (71 women, mean age=31.46, *SD*=10.02) completed a questionnaire using Amazon's Mechanical Turk.

Procedure. All participants received a 'neutral' case in which the chairman implements a program that will have the side effect of shifting the consumer base to be slightly older. Next, we manipulated which alternatives possibilities were relevant. In one condition, participants were asked to imagine possibilities in which the chairman had *wanted to avoid* shifting the consumer base, while in

the other condition, participants were asked to imagine possibilities in which the chairman had specifically *wanted* to shift the consumer base.

Participants were then asked whether they agreed or disagreed with the claim that the chairman intentionally shifted the consumer base. Participants rated this statement on a scale from 1 ('completely disagree') to 7 ('completely agree').

Results

Mirroring the effect of morality in Study 4a, participants who were asked to imagine possibilities in which the chairman had wanted to avoid the side-effect gave higher intentional action ratings (M=3.47, SD=2.14) than did those who imagined possibilities in which he had wanted to bring about the side-effect (M=2.83, SD=1.80), t(225.95)=-2.47, p=.014, d=.32, corrected for unequal variance.

Study 4a-b Discussion

In Study 4a we found that the impact of morality on judgments of intentional action were mediated by changes in participants' judgments of the relevance of alternative possibilities. In addition, Study 4b demonstrated that when participants were asked to consider an alternative possibility in which the agent wanted to avoid the side-effect, they judged that he acted more intentionally, mirroring the original effect of morality.

Study 5:

On the hypothesis we have been exploring, the central factor that affects judgments across these domains is not just whether an alternative possibility is considered, but whether that possibility is regarded as *relevant*. Thus, our account predicts that even when people are instructed to explicitly consider a particular possibility, they will have different judgments depending on whether or not they regard that possibility as relevant.

As demonstrated in Studies 1a, 2a, 3a, and 4a, participants tend to regard alternative possibilities that are morally bad as less relevant than those that are morally good. As a consequence, our account predicts that even if all participants are instructed to consider some alternative possibility, it should matter whether that possibility is morally good (and thus more relevant) or morally bad (and thus less relevant). We test this prediction in a final study.

Method

Participants. Two hundred and fifty participants (114 women, mean age=34.29, *SD*=10.88) completed a questionnaire using Amazon's Mechanical Turk.

Procedure. All participants read about a busy and distracted student athlete who failed a professor's chemistry class during a semester when the professor did not devote a great deal of time to teaching:

> Martin was a star football player in college who needed to complete a science class in order to continue to be eligible to play. He decided to enroll in Professor Smith's chemistry class.
>
> During that semester, Professor Smith gave the same lectures he'd been giving for ten years, and held a single review session before the final exam. Martin was sometimes tired in class and was distracted from the material and he was often busy after class and didn't have the time or energy to study on his own.
>
> By the end of the semester, Martin had earned a 48 on the midterm exam, a 65 in participation and a 55 on the final exam. At the end of the semester, Martin learned that he had failed the class. The athletics department was upset because Martin would not be eligible to play next year.

In the Relevant Possibility condition, participants were asked to consider and describe what would have happened if Professor Smith had instead decided to spend more time making his lectures easier to understand and held additional office hours for students who weren't understanding the material. In the Irrelevant Possibility condition, participants were asked to consider and describe what would have happened if Professor Smith had instead decided to alter Martin's grades at the end of the semester so that Martin (but no other student) received an extra 30 points on each of the

assignments.

After writing about what would have happened if these alternative possibilities had occurred, all participants were asked whether they agreed or disagreed with the statement that Martin's failing of the chemistry class was caused in part by Professor Smith. Participants rated this statement on a scale from 1 ('completely disagree') to 7 ('completely agree'). Lastly, demographic information was assessed. All stimuli, data, and analyses for this and all previous studies are available at: https://github.com/phillipsjs/RoA

Results

Participants who considered the morally good (and thus more relevant) alternative possibility were more likely to agree that Martin's failing was caused in part by Professor Smith (M=2.36, SD=1.39) than participants who considered a morally bad (and thus less relevant) alternative possibility (M=1.50, SD=0.84), t(188.00)=-5.85, p<.001, d=.76, corrected for unequal variance.

To ensure that this predicted difference did not arise simply because participants believed the relevant alternative action was actually more likely to lead to Martin passing the class, we also coded participants' written descriptions of what would have happened if Professor had acted differently (the full set of participants' responses can be found at https://github.com/phillipsjs/RoA). A comparison of participants' descriptions of what would have happened in these alternative possibilities revealed that participants directly mentioned that Martin would have passed the class more often when considering the irrelevant possibility (102 out of 132), than when considering the relevant possibility (69 out of 118), $X^2(1) = 9.33$, $p = 0.002$, $V = 0.193$. Moreover, even when we look only at responses from participants who explicitly mentioned Martin passing the class, we reconfirm the finding that Professor Smith is held to be more causal when participants considered a relevant alternative action (M=2.87, SD=1.49) rather than an irrelevant one (M=1.47, SD=0.85), t(98.05)=-7.04, p<.001, d=1.21, corrected for unequal variance.

Discussion

As confirmed by their written descriptions, participants in both conditions explicitly considered alternative possibilities that would have prevented the outcome from occurring. However, we observed the usual effect on participants' causal judgments when the alternative possibility they considered was a relevant one. Accordingly, the effect we observed between the two conditions appears to be due not merely to whether participants considered some alternative possibility, but instead to whether they regarded that possibility as relevant.

General Discussion

In the past decade, research has begun to uncover the surprising ways that people's moral judgments shape their non-moral cognition across a diverse array of domains (Cushman, Knobe & Sinnott-Armstrong, 2008; Knobe, 2003; Hitchcock & Knobe, 2009; Phillips & Young, 2011). While previous work has sought to account for the individual effect of morality in each domain (Machery, 2008; Menzies, 2010; Nichols & Ulatowski, 2007; Sytsma, Livengood, & Rose, 2012), the present studies sought to offer a general framework that explains the effects observed across all of these domains. Specifically, we propose that morality's influence in these diverse domains arises because morality directly affects whether people hold the alternatives to the agent's action to be *relevant*. Though such alternative possibilities were previously posited to play a role in these domains in theoretical work (Lewis, 1979; Bennett, 1995; Knobe, 2010; Aristotle, 350 BC/2002), the current work is the first empirical demonstration that morality's impact in all of these domains can be explained, at least in part, by changes in the relevance of alternative possibilities.

Using the materials that originally explored the effect of morality in the domains of freedom, causation, doing/allowing, and intentional action, Studies 1a, 2a, 3a, and 4a provided support for the idea that morally bad (vs. morally neutral) actions lead to different judgments about the relevance of

alternative possibilities and that these judgments of relevance partially mediated morality's effect on judgments across these domains. Studies 1b, 2b, 3b, and 4b offered more direct support for the relationship between the relevance of alternatives and judgments in each domain. Independent of morality, the relevance of alternative possibilities to the agent's actions showed the same pattern of influence that morality has been shown to have: participants judged that the actor was less forced, more of a cause, more did something rather than allowed it to happen, and acted more intentionally when certain alternatives were perceived as relevant. Finally, Study 5 provided evidence that the effect observed in these judgments is due to the perceived *relevance* of alternative possibilities rather than the mere consideration of them.

The role of the relevance of alternative possibilities

As we noted in the introduction, theoretical research in a number of fields (philosophy, linguistics, computer science, etc.) has converged on the central importance of providing a role for the representation of alternative possibilities. Moreover, recent work in these fields has emphasized the need to sort or rank alternative possibilities in terms of their relevance (Kratzer, 2012; Halpern & Hitchcock, 2014; Bello, in press). Here, we provide the first empirical evidence that this theoretical suggestion can help explain the impact of moral judgments across a number of various non-moral domains.

At the core of our account is a claim about the impact of moral judgments on intuitions about the relevance of alternative possibilities. More specifically, we argued that people show a general tendency to regard alternative possibilities as more relevant to the extent that they involve replacing morally bad things in the actual world with morally good alternatives. This tendency was observed in people's intuitions about each of the separate vignettes (Studies 1-4), and we propose that it explains the effects observed in all four domains.

If this theory is correct, it may also shed like on the question of *which* moral judgment it is that affects judgments across these domains. Previous research had assumed that it was a moral judgment that was in some way concerned with what the agent actually did, and the controversy has been over precisely what sort of moral judgment it is (Alicke, 2008; Phelan & Sarkissian, 2008; Nadelhoffer, 2004). What the current research suggests is that another relevant moral judgment is one that is about alternative possibilities. This is made most clear in Study 5, where the actual thing the professor did was exactly the same in both conditions, and what changed was only the moral valence of the alternative possibility that participants considered.

Though this general tendency can be found in all four domains, the precise details vary from one domain to the next. In each case, we find the same impact of moral judgment on the relevance of alternative possibilities, but these possibilities play importantly different roles in the different domains. Thus, to understand the role that alternative possibilities will play in a given domain, it is critical to consider existing theories of that specific domain.

In the domain of freedom and force, for example, we turned to existing philosophical accounts according to which an agent only performed an action freely if he or she had the possibility of behaving otherwise (Carr, 1988; Yaffe, 2003). Based on these accounts, we suggested that people's judgments about whether an agent *freely* performs some action depend in part on how they think about alternative possibilities in which the agent does *not* perform this action. The key claim here is that even in cases where it is obviously physically possible for the agent not to perform the action, a question arises as to whether not performing the action should be seen as a real possibility or as something so irrelevant that it should not be regarded as a genuine possibility at all. People's moral judgments impact their intuitions about the relevance of these possibilities and thereby impact their judgments as to whether the agent acted freely.

By contrast, in the domain of intentional action, we turned to existing theories according to

which intentional action judgments are based in large part on perceptions of the agent's *attitude*. We were therefore concerned with alternative possibilities concerning other attitudes that agent could have held. The key claim was that the agent's actual attitude is the same in both conditions (complete indifference to the outcome) but that this attitude is construed differently depending on which alternative attitudes we take as a comparison. If we compare the agent's actual attitude to possibilities in which he actively wants to bring about the outcome, the agent appears to be relatively uninterested in trying to bring it about. By contrast, if we compare it to possibilities in which he actively wants to avoid bringing about the outcome, the agent appears to be relatively interested in bringing it about. People's moral judgments impact which of these two kinds of possibilities seems most relevant, and thereby impact their judgments as to whether the agent acted intentionally.

Within the domains of causation and the doing/allowing distinction, there is a bit more controversy, with many theories positing an important role for alternative possibilities but different theories describing this role in quite different terms (Bennett 1993; 1995; Lewis, 1973). For present purposes, the key point is just that a broad spectrum of different theories would converge on the same basic prediction. Specifically, to the extent that people focus on possibilities in which an event does not occur, they should be more inclined to regard that event as a cause and as an instance of doing rather than allowing (for discussion, see Halpern & Hitchcock, 2014; Knobe & Szabó, 2013; Schaffer & Blanchard, in press).

Further research could continue to explore these effects at multiple levels. At one level, we need to examine the very general ways in which moral judgments can impact the representation of alternative possibilities. On another, we need to look separately at each domain and try to work out the precise role that possibilities play within that domain.

Situating the relevance of alternatives

Previous research has shown that morality can affect when and how people engage in

explicit counterfactual reasoning. Specifically, this research found that when people are asked what would have prevented negative outcomes from occurring, they tended to consider counterfactual alternatives to morally bad (as opposed to morally good) actions (N'gbala & Branscombe 1995; McCloy & Byrne 2000). The present studies, by contrast, did not consider when participants engaged in counterfactual reasoning or which counterfactuals they consider. Instead, participants were presented with specific alternative possibilities and were then asked to assess the degree to which these possibilities were relevant.

The present results thereby suggest that the impact of moral judgment is not just on which counterfactuals people consider in the first place. Rather, even when people are specifically instructed to consider a counterfactual, their moral judgments can impact their intuitions about whether the counterfactual is relevant or irrelevant. As demonstrated most clearly in Study 5, to the extent that a counterfactual is regarded as sufficiently irrelevant, people tend to feel that it does not truly represent a genuine possibility at all. It is this tendency, we argue, that lies at the root of the effects observed in all four domains.

While the positive account offered here is ambitious in that it attempts to offer a unified explanation of a number of diverse phenomena, it is also quite modest in two important respects. First, while we explained morality's impact on four different domains by demonstrating the role of the relevance of alternatives, we are not making the more general claim that *all* of morality's influence on non-moral judgments can be explained in this way. In fact, we expect that there are many domains in which alternative possibilities are unlikely to play any role at all, and in such cases, the account offered here simply cannot be appealed to in explaining the impact of morality. To take just one example, Ames and Fiske (2013) recently demonstrated that the moral status of an action affects people's assessments of the amount of monetary damage caused by that harm. Specifically, while participants in both conditions were told that the exact same amount of monetary expenses

were caused by the harm, participants in a condition in which the agent appeared to be more blameworthy later judged the total amount of damage (in dollars) to be greater. At present, we know of no research suggesting that alternative possibilities play a critical role in basic addition, and we find it highly unlikely that the current proposal will be able to be extended to effects of this nature. In addition to this first way that the scope of the current proposal is restricted, we also do not mean to suggest that morality's influence in the four domains we examined (freedom, causation, doing/allowing, and intentional action) is *solely* explained by morality's impact on the relevance of alternatives. After all, we only found that the relevance of alternatives partially mediated the effect of morality in these domains. Accordingly, it is likely that in each of these separate domains, there is some portion of the variance that is best accounted for by factors that are specific to that domain. Thus, we would argue that the current proposal is best understood as complementing previous research that may capture some domain-specific effects (for examples in the domain of intentional action, see, e.g., Sloman, Fernbach & Ewing, 2012; Sripada, 2010; Uttich & Lombrozo, 2010).

It is also likely that there are domain-general factors that affect these judgments. To take one example, research has suggested that individual differences such as gender or personality have an effect here (Feltz, 2007; Cokely & Feltz, 2009). Moreover, several theories have also proposed that morality may have a quite general influence on participants' responses. For example, it has been suggested that people's moral judgments can impact their responses by affecting the conversational pragmatics (Adams & Steadman, 2004a; 2004b; Driver, 2008a; 2008b) or by triggering a motivation to justify the claim that the agent is blameworthy (Alicke, 2000; Ditto, Pizarro, & Tannenbaum, 2009). This previous research has demonstrated a pervasive phenomenon whereby people's understanding of the events that actually occurred is altered in ways that support and justify their desire to blame an agent. It is certainly possible that some portion of morality's influence will be best explained by some of these pervasive phenomena. As Cushman (2014) has argued, it is often the

case that big effects are best explained by a combination of many separate smaller effects.

Future work and conclusion

Across five studies, we found support for the idea that morality's influence in the domains of freedom, causation, doing/allowing, and intentional action can be explained, at least in part, by changes in judgments of the perceived relevance of alternative possibilities. Studies 1a, 2a, 3a and 4a demonstrate that these judgments of relevance mediate morality's influence in four domains. Studies 1b, 2b, 3b and 4b more directly test the role of alternative possibilities in these four domains by demonstrating that non-moral changes in the relevance of alternative possibilities have a similar effect in each of these domains.

The current proposal suggests a number of promising avenues for further research. First, future work should explore whether the present account can be extended to other domains in which morality has been shown to have an impact. Existing work has found an impact of morality not only in the four domains explored here but also on judgments of knowledge, happiness and innateness, among others (Beebe & Buckwalter, 2010; Phillips, Misenheimer & Knobe, 2011; Phillips, Nyholm & Liao, in press; Knobe & Samuels, 2013). Future work should examine judgments in these other domains, perhaps using the same methods employed here.

Second, the present account suggests that judgments in the four domains explored here are influenced by morality because alternative possibilities play a central role in each of these domains. Thus, our account predicts that morality should also have an influence in any additional domains in which alternative possibilities play a central role. Future research could look to other domains in which possibilities play a role and ask whether an impact of morality can be found there as well.

Third, Studies 1b, 2b, 3b and 4b showed that the relevance of alternative possibilities plays a role in these judgments even when the moral valence of an action is held constant. This suggests that *other* factors that influence judgments of the relevance of alternatives should also have an effect

in each of these domains. For example, existing theoretical work suggests that the relevance of alternative possibilities should not only be affected by moral judgments but also by judgments about frequencies (Halpern & Hitchcock, 2014; Kahneman & Tversky, 1982; Kratzer, 2012; Schaffer & Blanchard, in press). The present account therefore predicts that facts about frequencies should impact judgments in all four domains, and should do so in a way that parallels the present findings regarding the impact of morality.

In short, the present work offers a unifying framework focused on the relevance of alternative possibilities that can help explain morality's often puzzling influence on non-moral judgments, and opens up a number of exciting avenues for future research.

References

Adams, F., & Steadman, A. (2004a). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis*, 64, 173-181.

Adams, F., & Steadman, A. (2004b). Intentional actions and moral considerations: Still pragmatic. *Analysis, 64,* 268-276.

Alicke, M. D. (2008). Blaming badly. *Journal of Cognition and Culture*, 8(1), 179-186.

Alicke, M.D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126(4), 556-574.

Ames, D.L., & Fiske, S.T. (2013). Intentional harms are worse, even when they're not. *Psychological Science, 24(9),* 1755-1762.

Aquinas, T. (1273/1920). *The Summa Theologica of St. Thomas Aquinas*, Second and Revised Edition. Translated by Fathers of the English Dominican Province.

Aristotle (350 BC/ 2002). *Nicomachean Ethics*. (Rowe & Broadie, Trans.). Oxford University Press.

Barry, C., Lindauer, M. & Øverland, G. (in press) Doing, Allowing, and Enabling Harm: An Empirical Investigation. In J. Knobe, T. Lombrozo & S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy*. Oxford: Oxford University Press.

Beebe, J. R., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language*, *25*(4), 474-498.

Bello, P. (in press). Machine ethics and modal psychology. *Synthese Library*.

Bennett, J. (1993). Negation and abstention: Two theories of allowing. *Ethics*, 104: 75–96.

Bennett, J. (1995). *The act itself*. Oxford: Clarendon Press.

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.

Berlin, I. (1969/2002). Two Concepts of Liberty. In I. Berlin, *Four Essays on Liberty*, London: Oxford University Press.

Byrne, R. M. (2005). *The rational imagination: How people create alternatives to reality*. MIT Press.

Carr, C. L. (1988). Coercion and freedom. *American Philosophical Quarterly*, 25, 59-67.

Cokely, E. T., & Feltz, A. (2009). Individual differences, judgment biases, and theory-of-mind: Deconstructing the intentional action side effect asymmetry. *Journal of Research in Personality*, 43(1), 18–24.

Copley, B. & Wolff, P. (2014). Theories of causation can and should inform linguistic theory. In. B. Copley, F. Martin, & N. Duffield (Eds.), *Forces in grammatical structures: Causation between linguistics and philosophy*. Oxford University Press.

Cova, F. (2014). Frankfurt-style cases User Manual: Why Frankfurt-style enabling cases do not necessitate tech support. *Ethical Theory and Moral Practice*, 17(3), 505-521.

Cova, F., & Egré, P. (in press). Moral asymmetries and the semantics of "many." *Semantics and Pragmatics*.

Cova, F. & Kitano, Y. (2014). Experimental philosophy and the compatibility of free will and determinism: a survey. *Annals of the Japan Association for Philosophy of Science*, 22, 17-37.

Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition, 108*, 281-289.

Cushman, F. (2014). What scientific idea is ready for retirement? Big Effects Have Big Explanations. Retrieved from http://edge.org/response-detail/25508.

Ditto, P.H., Pizarro, D.A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D.M. Bartels, C.W. Bauman, L.J. Skitka, & D.L. Medin (Eds.), *Moral judgment and decision making* (pp. 307-338). San Diego, CA: Academic Press.

Donagan, Alan, 1977, *The Theory of Morality*, Chicago: The University of Chicago Press.

Driver, J. (2008a). Attributions of causation and moral responsibility. In W. Sinnott-Amstrong (Ed.) *Moral psychology (Vol. 2):, The cognitive science of morality: Intuition and diversity* (pp. 423-439).

Cambridge, MA: MIT Press.

Driver, J. (2008b). Kinds of norms and legal causation: Reply to Knobe and Fraser and Deigh, in W. Sinnott-Amstrong (Ed.) *Moral psychology (Vol. 2):, The cognitive science of morality: Intuition and diversity* (pp. 459-461). Cambridge, MA: MIT Press.

Epstude, K., & Roese, N. J. (2008). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Personality and Social Psychology Review*, 12(2), 168–92.

Feltz, A. (2007). An anomaly in intentional action ascription : More evidence of folk diversity. *Proceedings of the 29th Annual Cognitive Science Society*, 1748.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2015). How, whether, why : Causal judgments as counterfactual contrasts. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.

Halpern, J., & Hitchcock, C. (2014). Graded Causation and Defaults. *British Journal for the Philosophy of Science*.

Hitchcock, C. & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 11,* 587-612.

Hindriks, F. (2014). Normativity in Action: How to Explain the Knobe Effect and its Relatives. *Mind & Language*, *29*(1), 51-72.

Kagan, Shelly, 1989, *The Limits of Morality*, Oxford: Oxford University Press.

Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136-153.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, 201–208. New York: Cambridge University Press.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33,* 315-329.

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

Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral Psychology.* Cambridge, MA: MIT Press.

Knobe, J., & Samuels, R. (2013). Thinking like a scientist: Innateness as a case study. *Cognition*, *126*(1), 72-86.

Knobe, J., & Szabó, Z. G. (2013). Modals with a test of the deontic. *Semantics & Pragmatics*, *6*.

Kratzer, A. (2012). *Modals and conditionals.* Oxford: Oxford University Press.

Lagnado, D. a, Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, 37(6), 1036-1073.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70(17), 556-567.

Lombrozo, T. (2010). Causal-Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions. *Cognitive Psychology*, 61(4), 303–32.

Machery, E. (2008). The folk concept of intentional action: Philosophical and experimental issues. *Mind & Language, 23,* 165-189.

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.

McCloy, R., & Byrne, R. (2000). Counterfactual thinking about controllable events. *Memory and Cognition, 28,* 1071-1078.

Menzies, P. (2010). Norms, causes, and alternative possibilities. *Behavioral and Brain Sciences,* 33, 346-347.

Miller, J. & Feltz, A. (2011). Frankfurt and the folk: An empirical investigation. *Consciousness and Cognition*, 20, 401-414.

N'gbala, A., & Branscombe, N.R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology, 31,*139-162.

Nadelhoffer, T. (2004). Blame, badness, and intentional Action: A Reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology*, 24(2), 259-269.

Nduibuisi, B. & Byrne, R.M.J. (2013). Intentionality and choice. In M. Knauff, M., Pauen, N., Sebanz, & Wachsmuth, I. (Eds.) *Proceedings of the 35th Annual Conference of the Cognitive Science Society,* 1970-1975. Austin TX: Cognitive Science Society.

Nichols, S., & Ulatowski, J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language, 22(4),* 346-365.

Pearl, J. (2000). *Causality: models, reasoning and inference* (Vol. 29). Cambridge: MIT press.

Phelan, M. T., & 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.

Phillips, J., & Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry*, 20, 30-36.

Phillips, J., Misenheimer, L., & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review, 3*(3), 320-322.

Phillips, J., Nyholm, S. & Liao, S. (in press). The good in happiness. In J. Knobe, T. Lombrozo & S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy*. Oxford: Oxford University Press.

Portner, P. (2009). *Modality*. Oxford: Oxford University Press.

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.

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

Schaffer, J., & Blanchard, T. (in press). Cause without default. In H. Beebee, C. Hitchcock, & H. Price (Eds.), *Making a Difference*. Oxford: Oxford University Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed

flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Sloman, S. a., Fernbach, P. M., & Ewing, S. (2012). A Causal Model of Intentionality Judgment. *Mind & Language*, 27(2), 154–180.

Spencer, S.J., Zanna, M.P., & Fong, G.T. (2005). Establishing a causal chain: Why experiments are often more effective in examining psychological process than meditational analyses. *Journal of Personality and Social Psychology, 89,* 845-851.

Sripada, C. S. (2010). The Deep Self Model and asymmetries in folk judgments about intentional action. *Philosophical Studies*, 151(2), 159–176.

Strickland, B., & Suben, A. (2012). Experimenter philosophy: The problem of experimenter bias in experimental philosophy. *Review of Philosophy and Psychology*, *3*(3), 457-467.

Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43(3),* 814-820.

Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition, 116(1),* 87-100.

Walsh, C. R., & Sloman, S. a. (2011). The Meaning of Cause and Prevent: The Role of Causal Mechanism. *Mind and Language*, 26(1), 21–52.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, 56(2), 161-169.

Woolfolk, R., Doris, J. & Darley, J. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100, 283-301.

Woodward, J. (2004). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Yaffe, G. (2003). Indoctrination, coercion and freedom of will. *Philosophy and Phenomenological Research*, *67*(2), 335-356.

Young, L., & Phillips, J. (2011). The paradox of moral focus. *Cognition,* 119, 166-178.

[1] A direct replication of this study using a within-subjects design yielded in a similar pattern of responses. Replicating previous results, participants rated Professor Smith as more of a cause (M=5.17, SD=1.63) than the administrative assistant (M=2.49, SD=1.62), $t(399)=20.31$, $p<.001$, d=1.65. Additionally, participants regarded the possibility in which the Professor did not take a pen

to be more relevant (M=4.86, SD=1.82) than the possibility in which the administrative assistant did not take a pen (M=2.91, SD=1.90), $t(398)=10.47$, $p<.001$, d=1.05 (reverse coded).