ATTRIBUTIONS OF MENTAL STATE CONTROL:

CAUSES AND CONSEQUENCES

Corey Cusimano

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

____

Geoffrey Goodwin, PhD, Professor of Psychology

Graduate Group Chairperson

____

John Trueswell, PhD, Professor of Psychology

Dissertation Committee:

Paul Rozin, PhD, Professor of Psychology

Jonathan Baron, PhD, Professor of Psychology

ACKNOWLEDGMENT

Many people helped me shape the arguments and experiments that make up this dissertation. I thank Coren Apicella, Jon Baron, Geoff Goodwin, Bertram Malle, Andrew Monroe, Olivia Podos, Ike Silver, Joe Simmons, Kris Smith, Jessie Sun, and Dan Swingley, for their help conducting this research and their thoughtful feedback on it.

I am indebted to my advisor, Geoff Goodwin, more than is possible to briefly say here. I remain in disbelief that you have been able to accomplish anything else given the incredible amount of attention you provided me these past four years. The feedback you provided, and the bar you set for me, raised my capacity and confidence as a scientist and writer more than I considered possible. But my greatest debt to you – more than the opportunity to pursue a PhD, more than the opportunity to freely pursue my own wild research ideas, and more than the years of helpful and insightful advice – is the template you provide for being a teacher and psychologist. You are one of the most well-rounded, responsible, professional, and thoughtful people I have ever known, and this fact means that you will still be pushing me to improve long after I have left Solomon Hall.

Ed Royzman has become a dear friend and a valuable collaborator. I am a far better psychologist for having known you. But more than that, you have been a source of humor and joy these past few years, and your encouragement and support helped me during my darkest times. I look forward to many more years of collaboration.

I have only recently become close with Paul Rozin but, even in this short time, he has left a lasting impression. This past year has been one of the best in my life and that is in large part due to your wonderful kindness, enthusiasm for teaching, and infectious love

of the world. You remind me why I pursued this career in the first place and inspire me to pass this joy along to as many people as I can.

I owe a special dept to Sudeep Bhatia, Jon Baron, and Dan Swingley for graciously advising me at various points through my PhD. I constantly think about the many nuggets of wisdom you all passed along to me over the years. I owe a special thanks to Jon Baron, who in addition to being generous, responsive, and wise, was unique in his extreme and continued enthusiasm for the topics I pursued in graduate school.

Pat Spann, Yuni Thorton, Paul Newlon, and Sara Jaffee were critical to my survival these past four years. They all looked after us and helped me out more times than I can recall. Pat Spann in particular worked harder than anyone else to turn Solomon Hall into a place I wanted to be every day. We are all lucky to have you.

Kelly Allred, Trevor Brothers, Neal Fox, Julia Franckh, Hannah Hastings, Jack Keefe, Joanna Korman, Anna Leshingskaya, Sahil Luthra, Andrew Monroe, Josiah Nunziato, Aru Sarin, Rachel Schwartz, Ike Silver, Valerie Snow, Hanne Watkins, Cliff Workman, and Lisa Yankowitz. You all made the past four years worthwhile. Thinking now of all you have collectively done for me makes me embarrassed of how lucky I am. I would not have made it without you.

I owe a huge debt to my family for their support, advice, and inspiration. Thank you to my many parents, who have all worked so hard to give me a good life.

And finally, to my past teachers, Joe Hoffman, Josh Knobe, Geoff Sayre-McCord, Dorit Bar-On, and Bertram Malle, I am here because you saw promise in me. Thank you.

ABSTRACT

ATTRIBUTIONS OF MENTAL STATE CONTROL: CAUSES AND CONSEQUENCES

Corey Cusimano

Geoffrey P. Goodwin

A popular thesis in psychology holds that ordinary people judge others' mental states to be uncontrollable, unintentional, or otherwise involuntary. The present research challenges this thesis and documents how attributions of mental state control affect social decision making, predict policy preferences, and fuel conflict in close relationships. In Chapter 1, I show that lay people by-and-large attribute intentional control to others over their mental states. Additionally, I provide causal evidence that these attributions of control predict judgments of responsibility as well as decisions to confront and reprimand someone for having an objectionable attitude. By overturning a common misconception about how people evaluate mental states, these findings help resolve a long-standing debate about the lay concept of moral responsibility. In Chapter 2, I extend these findings to interpersonal emotion regulation in order to predict how observers react to close others who experience stress, anxiety, or distress. Across six studies, I show that people's emotional support hinges on attributions of emotion control: People are more inclined to react supportively when they judge that the target individual cannot regulate their own emotions, but react unsupportively, sometimes evincing an intention to make others feel bad for their emotions, when they judge that those others can regulate their negative emotion away themselves. People evaluate others' emotion control based on assessments of their own emotion regulation capacity, how readily reappraised the target's emotion is, and how rational the target is. Finally, I show that judgments of emotion control predict self-reported supportive thoughts and behaviors in close relationships as well as preferences for university policies addressing microaggressions. Lastly, in Chapter 3, I show that people believe that others have more control over their beliefs than they themselves do. This discrepancy arises because, even though people conceptualize beliefs as controllable, they tend to experience the beliefs they hold as outside their control. When reasoning about others, people fail to generalize this experience to others and instead rely on their conceptualization of belief as controllable. In light of Chapters 1 and 2, I discuss how this discrepancy may explain why ideological disagreements are so difficult to resolve.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

The Puzzle of Mental State Responsibility

We are not idle observers of others. When those around us cause harm, or act inappropriately, we react. Sometimes we tell others what we saw, sometimes we break off our relationship with the offender or urge others to do the same. And sometimes we hold the offender *morally responsible* for her behavior. Unlike other types of reactions, holding someone responsible often means imposing costs on them, including expressing anger toward them, punishing them, criticizing them, demanding that they explain themselves or make amends, or otherwise making them feel bad (Coates & Tognazzini, 2013; Malle, Guglielmo, & Monroe, 2014). These behaviors teach them that what they did was unacceptable, communicate our expectations moving forward, and provide motivation to the transgressor to fulfill those expectations (Cushman, 2013; Malle et al., 2014; Heider, 1958).

However, the costs imposed by moral responsibility can sometimes be extreme, such as permanent injury, long-term incarceration, or death. Even mild forms of being held responsible, such as being criticized or reprimanded, are highly unpleasant to experience. For these reasons, it is typically only acceptable to criticize someone, confront them, or make demands of them if they *deserve* it (Alicke, 2000; Coates & Tognazzini, 2013; Sabini & Silver, 1998; Shaver, 1985). Indeed, punishing or blaming someone who does not meet the standard required is itself deemed blameworthy (Malle et al., 2014). This observation raises the question: what causes people to believe someone deserves to be held responsible?

A great deal of work in the past fifty years suggest that what it takes to hold someone responsible for something is for the individual in question to have had control over it (Alicke, 2000; Fincham & Jaspers, 1980; Malle et al., 2014; Shultz, Schleifer, & Altman, 1981; Weiner, 1995). Indeed, recent work shows that, when faced with some moral violation, people seek out information about whether that person was the cause of the bad outcome, how foreseeable that outcome was, and whether they could have prevented or avoided it (Guglielmo & Malle, 2017). When people learn that someone lacked the proper control, they adjust their blame down (Monroe & Malle, 2018). For instance, people judge others as more responsible for causing harm intentionally, rather than accidentally, in part because of the additional control they exert over the outcome (Alicke, 2000; Lagnado & Channon, 2008). Likewise, people hold others responsible for preventable outcomes more than unpreventable ones. For instance, people blame someone for failing an exam when that person could have passed (e.g., because they chose not to study even though they could have studied) rather than when they did not have the ability to pass (e.g., they are unintelligent; Weiner, 1995). Control affects whether people blame each other for their illnesses, too (Haslam & Kvaale, 2015). For instance, when people learn that some stigmatizing illness, like depression or schizophrenia, is not controllable (for instance, because it is caused by someone’s genes rather than their behavior) people hold the sick individual less responsible for it (Lebowitz & Ahn, 2014). Because the connection between control and moral responsibility has replicated across so many contexts, it appears to be an essential feature of moral responsibility.

Like other kinds of conduct, people hold each other responsible for their emotions, desires, and beliefs. Atheists living in religious communities are verbally harassed and pressured to renounce their beliefs (Hammer, Cragun, Hwang, & Smith, 2012). Similarly, highly religious people express outrage towards others who hold attitudes that violate religious taboo, such as holding heretical beliefs or failing to hold other prescribed attitudes, such as when someone feels disrespect towards one's parents (Cohen & Rozin, 2001; Tetlock et al., 2000). People want to punish those who experience feelings of *schadenfreude* (Gromet, Goodwin, & Goodman, 2016) or act in a way that reveals a "wicked" desire (e.g., hoping that harm befalls another person; Inbar, Pizarro, & Cushman, 2012). People confront and punish close friends, romantic partners, or family members who hurt their feelings by disrespecting, disliking, or not caring about them (Leary et al., 1998). And, in some instances, mental state punishment has been codified into law. For instance, in some states in the US, people support or have passed laws that punish perpetrators who are motivated by bias while committing crimes, thereby increasing their prison time or fines relative to others who commit otherwise identical crimes (i.e., hate crimes). In sum, just like ordinary behavior, people believe others' mental states can violate social or moral norms and will hold them responsible.

And yet, unlike ordinary behavior, lay people appear to judge mental states as, by-and-large, outside people's control. At least, this appears to be the dominant position amongst psychologists and other scholars. Expressing this view, Gilovich and Regan (1986) wrote that mental states "do not necessarily involve any choice on the part of the person from among alternatives; *they just happen*" (p. 349, emphasis original). In a similar vein, Malle and Knobe (1997a) assume that, to ordinary people, "prototypical actions . . . are both intentional and observable, whereas prototypical experiences (e.g. ‘Ben is excited’) are both *unintentional* and unobservable” (p. 289; emphasis added). In reference to immoral mental states, Adams (1985) stipulates that “jealousy, hatred, and other sorts of malice; contempt for other people, and the lack of a hearty concern for their welfare; or in more general terms, morally objectionable states of mind, including corrupt beliefs as well as wrong desires” are “involuntary” moral transgressions (p. 4). This assumption about how ordinary people judge mental states is held by many others (e.g., D’Andrade, 1987; Katz & Postal, 1964; Sabini & Silver, 1998; Smith, A., 2005; Smith, H., 2010). Because people apparently view mental states as uncontrollable, many scholars view the ordinary practice of mental state blame as a devastating counter example to the classic view that moral responsibility requires control (Adams, 1985; Smith, A., 2008; Pizarro, Tannenbaum, & Uhlmann, 2012; but see Sankowski, 1977, for a contrary view). Indeed, according to Smith, A. (2008), mental state responsibility reveals to us that “in our day-to-day lives we simply take for granted that people are responsible for much more than what they voluntarily choose to do” (p. 87).

If not control, then what alternative standard do lay people apparently have in mind when they hold others responsible for their mental states? One promising alternative offered is that judgments of moral responsibility stem from evaluations of *moral character* – i.e., judgments of whether the individual is a *good* person or a *bad* person (Pizarro & Tannenbaum, 2011; see Bayles, 1982, for a philosophical articulation of this view). Critically, while control can influence perceptions of character (e.g., voluntary behaviors are more revealing of character than situationally induced behaviors, Jones & Davis, 1965; Monroe & Reeder, 2011), qualities or behaviors can reveal a person’s character absent control. For instance, if someone complains that their friend is “incapable of feeling compassion towards others,” then that person is indicting their character even while stipulating that the negative quality in question is outside of their control (c.f., Adams, 1985). Rather, what appears to be most relevant when evaluating someone’s character is that the conduct diagnoses some stable feature of that person’s psychology, including their core values, cognitive or emotional capacities, or morally-relevant dispositions of thought or behavior (c.f. Critcher, Inbar, & Pizarro, 2013; Reeder, 1993, 2009). Because mental states are often the output of stable features of an individual, such as their capacity or disposition for reason or empathy, and because certain elements of moral character are constituted by mental states (e.g., having general good or ill will towards others), mental states should be seen as highly revealing of character even if they are viewed as uncontrollable.

Consistent with a character account of moral responsibility, ordinary people treat emotions, beliefs, and desires as highly diagnostic of someone’s character. Emotions are considered especially revealing. For instance, even without corresponding prosocial or antisocial behavior, people will make inferences about someone’s character based solely on that person’s emotional reaction to someone else’s ill fortune. When someone does not feel upset when they witness someone else in pain, people attribute negative moral traits to that person (e.g., “callousness,” Szczurek, Monin, & Gross, 2012). And, if that person feels pleasure at another’s pain, then people judge that person to be evil (Gromet et al., 2016). Even when people act in otherwise prosocial ways, observers prioritize mental state information when making character inferences. For instance, people who help others are seen as having good character when their behavior signals prosocial attitudes and values (e.g., “this person cares about her friends”) rather than selfish or self-serving desires (Ames & Johar, 2009; Uhlmann, Zhu, & Tannenbaum, 2013).

Additionally, mental states, and the character evaluations they lead to, play a central role in determining when people form and dissolve relationships. When people select romantic partners, friends, and close associates, they prioritize shared values, morals, political viewpoints, and belief systems (Brandt, Reyna, Chambers, Crawford, & Wetherell, 2014; Gibbs, Ellison, & Heino, 2006; Haidt, Rosenberg, & Hom, 2003; Murray, Holmes, Bellavia, Griffin, & Dolderman, 2002; Skitka, Bauman, & Sargis, 2005; Sprecher & Hendrick, 2004). When people infer poor character or other stable mental states, for instance by having an inappropriate emotional reaction to something, people seek to avoid that individual (Ames & Johar, 2009; Szczurek, et al, 2012).

To summarize, according to control theories of moral responsibility (Figure 1.1A), control is necessary for holding someone responsible whereas other considerations, such as inferences of poor character are not sufficient (e.g., Cushman, 2015; Sabini & Silver, 1998). Mental state blame is supposedly a counterexample to this theory of moral judgment. This is because people appear to hold others responsible for their mental states, and do so while apparently believing that mental states are not controllable. That is, mental state responsibility appears to show support for an alternative theory, like the one shown in Figure 1.1B, in which inferences of poor character license decisions to blame, punish, or confront someone (in addition to licensing other reactions such as avoiding or gossiping about them).

But while mental states appear to present a challenge to classic, control-based, models of moral judgment, no work has actually empirically investigated whether people actually evaluate and react to others' immoral mental states this way. Specifically, there are two questions to ask: First, is it true that lay people believe mental states are outside people's control? And, second, is it true that lay people believe control is unnecessary to attribute responsibility to someone for a mental state? As I summarize below, past work is equivocal on these questions.

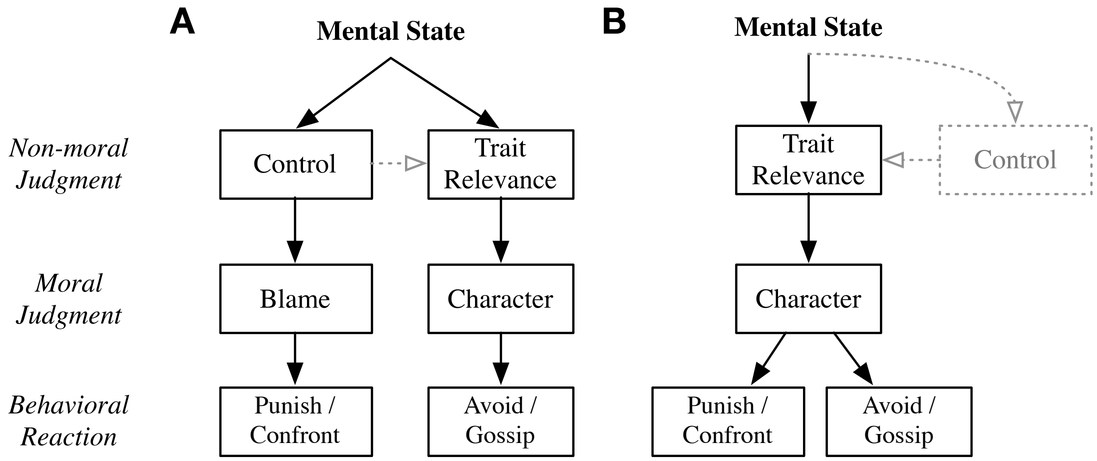

Figure 1.1. Two models of everyday moral judgment. (A) Holding someone morally responsible is distinct from other reactions to immoral conduct in that it requires control. (B) Holding someone responsible is determined by evaluations of their moral character, just like other reactions are. Control may influence character inference but is not necessary for moral responsibility.

Do lay people judge mental states as outside control?

If ordinary people believe mental states are outside people's control, then the everyday practice of holding each other responsible for our mental states would constitute a counter example to control theories of responsibility. But do people judge mental states to be uncontrollable? Little work has investigated this question and the work that has paints an unclear picture.

Some studies have suggested that people attribute a moderate degree of control to others over their mental states. For instance, Schlesinger (1992) gave people sentences containing mental state verbs (e.g., “A likes B” or “A impressed B”) and asked them to rate how much control (Studies 1– 4, 6) or intentionality (Study 5) the subject or object of the sentence (e.g., “A” or “B”) had over the event. Schlesinger’s aim was to test whether people tend to attribute agency to whoever (or whatever) is in the subject position of the sentence. This is exactly what he found: the subject of the sentence was routinely rated as having more agency over the event than the object. However, an auxiliary finding, one more relevant to our project, was that subjects in Schlesinger’s studies attributed moderate levels of control and intentionality to agents in both the subject *and* object positions of the sentence (where perceived control would not be inflated by syntactic cues). For instance, agents were judged to have middling control (4.69 of 9) over feeling “excited” even when they were in the object position of a sentence. Thus, these results suggest that people are willing to attribute at least some control to the experiencers of mental states.

Similarly, Turri, Rose, and Buckwalter (2018) provide evidence that people sometimes judge beliefs to be controllable. Across a series of studies, Turri et al. presented subjects with simple vignettes describing a person asserting their conscious decision to believe (or refuse to believe) something, for instance, that a legislative bill would pass, or that extraterrestrial life would be discovered (e.g., the person announces, “I want to continue as part of this administration, so I choose to believe the bill will pass”). They found that subjects later reported high agreement with statements recapitulating that agent’s decision (e.g., “Mrs. Platters can choose to believe the bill will pass”), and separately, with statements indicating that the target possesses the belief in question (e.g., “Because she made that choice, now Mrs. Platters believes that the bill will pass”). Because the subjects agreed with the follow-up statements, the authors concluded that that people judge it “conceptually possible” (p. 1) that someone could exert voluntary control over a belief. In two of these studies, they observed a similar finding for closely related mental states (e.g., holding opinions or having doubts) as well as some unrelated mental states (e.g., wanting, feeling excited, and intending).

These studies seem to show that people countenance the possibility of mental state control, but for our present purposes— assessing everyday judgments of mental state control—they are limited in several ways. Asking subjects to agree with a statement that recapitulates the earlier content of the vignettes may not necessarily capture people’s default expectations of others’ mental state control. For instance, if a character in a vignette announces that she just performed a backflip, people may be inclined to agree with a statement attributing to her this capacity, while still generally expecting that most people, most of the time, are not so capable. Additionally, the overall sampling of mental state contents in these studies was limited, raising doubts about whether subjects’ judgments would generalize to a wider range of everyday situations.

Indeed, several other studies have come to conclusions opposite those made by Schlesinger (1992) and Turri et al. (2018), suggesting instead that people judge mental states as passive and unintentional. Johnson, Robinson, and Mitchell (2004) conducted a study using methods similar to those used by Schlesinger. They found that people tended to judge actions such as "Sarah harasses Amy," on average to be easier to control than mental states such as "Sarah envies Amy." Additionally, they observed that, overall, mental states were judged to be on the "difficult to control" side of a 9-point scale ranging from 1: *probably very difficult* [to control] to 9: *probably very easy* [to control] (Study 1, $M_{\text{mental states}} = 3.92$; Study 2, $M_{\text{mental states}} = 4.63$; scale midpoint = 5).

Gilovich and Regan (1986) reported a study in which a variety of mental experiences, gathered from diary entries, were judged by their experiencers as driven more by situational factors than by dispositional factors. In contrast, the same subjects typically judged their own actions as driven more by dispositional than situational factors. Gilovich and Regan (1986) interpreted these data as suggesting that, whereas actions are voluntarily chosen, "many of our experiences 'happen' to us, with little or no exercise of choice or decision on our part" (p. 349). Consistent with this idea, independent judges rated subjects' actions as reliably more chosen than so-called experiential mental states, like feeling an emotion. However, there are limits to the generalizability of this study: it relied on a small set of diary entries ($N = 19$), and the overall number and type of mental states subjects recalled was not documented.

Malle and Knobe (1997b) conducted a study investigating lay attributions of the intentionality of a wide variety of ordinary behaviors. As a part of this study, they asked subjects to rate several mental state scenarios, including "Anne was in a great mood today," "Anne had a craving for cherries after dinner," and "Anne believed that she had the flu," on their degree of intentionality. Subjects rated these mental states as largely unintentional (average ratings were 2.70, 2.23, and 2.69, respectively, on a 1–7 scale). These data are suggestive, but because they are only based on three, potentially idiosyncratic items, they do not license a general conclusion about mental state control (which was not Malle and Knobe's aim).

Taken as a whole, the work summarized above leads to no clear picture of whether ordinary people conceptualize mental states as controllable and intentional, or not. Existing studies have relied on limited and ad hoc sampling of mental states, and they have yielded conflicting conclusions. Moreover, most of the studies reported above obtained control judgments by using highly artificial statements (e.g., "A feared B") or mental states completely divorced of context (e.g., "Sarah envies Amy"), raising the question of whether subjects' judgments generalize to real-life contexts. For these reasons, my first goal in this chapter was to test to what extent lay people genuinely view everyday mental states as controllable or uncontrollable.

Do lay people rely on attributions of control when reacting to mental states?

Although there is a great deal of work showing that people infer character from others' mental states (see above), there is comparatively less work testing whether people's reaction to others' mental states integrate attributions of control. One area that has received a great deal of attention, however, is how people react to *their own* mental states. Specifically, people's attributions of control seem to predict their motivations, strategies, and sense of responsibility with respect to their own emotions (see Ford & Gross, 2019, for a review). Several studies now show that individuals are inclined to engage in cognitive reappraisal to the extent that they think that their own emotions are controllable (e.g., De Castella et al., 2013; Ford, Lwi, Gentzler, Hankin, & Mauss, 2018; Kappes & Schikowski, 2013; Kneeland, Nolen-Hoeksema, Dovidio, & Gruber, 2016;

Schroder, Dawood, Yalch, Donnellan, & Moser, 2015). Additionally, people who judge that they ought to be able to control their emotions are more likely to get angry at themselves for episodes of unwanted emotionality (Mitmansgruber, Beck, Höfer, & Schüßler, 2009). If people evaluate others' mental states the way that they apparently evaluate their own, then we should expect control to play an integral role.

There is some evidence that people base their reaction to others' mental states on control as well. For instance, a seminal study demonstrated religious differences in the opprobrium directed toward holders of inappropriate mental states (Protestants being harsher judges than Jews), which were partially mediated by perceived control over the offending mental states (Cohen & Rozin, 2001). Other research has shown that attributing sexual orientation to personal choice (or upbringing) rather than biological predisposition predicts negative affective responses toward homosexuals, the belief that homosexuality is unacceptable, and opposition to equal rights for same sex couples (Haider-Markel & Joslyn, 2008). However, these studies are limited in two respects. First, they do not attempt to distinguish between moral responsibility and character, or measure whether one is more important to responsibility than the other. And second, they are correlational, leaving open the possibility that attributions of control are downstream of character assessments (c.f. Nadler & McDonnell, 2012). We address each of these limitations in the studies below.

The current studies

In sum, past work is equivocal regarding whether lay people attribute to others' any substantive degree of control over mental states, or whether people base their reactions to others' immoral mental states. We address this across three studies, reported below. In Study 1.1, we conducted an exploratory test of what kinds of mental states, if any, are treated as uncontrollable. Inspired by past work suggesting that different mental states may be treated as voluntary or involuntary (e.g., D'Andrade, 1987), we examined a wide range of mental state types. In Studies 2 and 3 we investigate how people reason about moral responsibility when evaluating others' immoral mental states. In Study 1.2, we measure perceived control of immoral emotions, desires, beliefs, and evaluations and ask subjects to make character and responsibility judgments. Finally, in Study 1.3, we manipulate the perceived controllability of immoral states and test whether that effects perceived responsibility or character, as well as responsibility-relevant or responsibility-irrelevant behaviors.

Study 1.1

Study 1.1 was an exploratory investigation of the degree of control people attribute to others over their everyday mental states. We first asked one sample of our target population (University of Pennsylvania undergraduates) to provide examples of everyday mental states. We then selected the most frequent examples and asked a separate sample from the same population to rate how much control others possess over each mental state. This procedure helped ensure that our results reflected everyday mental state reasoning (by drawing upon examples people commonly think about), while also minimizing experimenter bias.

To assess the degree of control that people attribute to different mental states, we compared subjects' ratings with observable behavior foils, including intentional acts (e.g., *talk, avoid*), accidents (e.g., *slip, fall*), and uncontrollable behaviors (e.g., *sneeze, shiver*). This strategy confers three main benefits. First, these foils anchor the rating scales across subjects and studies. Second, they act as checks that subjects are responding in a sensible way (e.g., unintentional behaviors should be judged at the floor of the scale, intentional behaviors at the ceiling). Third, these foils allow us to assess ratings of controllability against intuitively understood benchmarks. For instance, if a particular mental state is indistinguishable from intentional behaviors, we can infer that people typically regard it as fully controlled or intended, whereas if it is judged indistinguishably from uncontrolled or unintentional behaviors, we can infer that it is regarded as fully uncontrolled or unintentional.

Method

Stimulus generation and selection. Eighty University of Pennsylvania students participated (57 female) in a sentence completion task for course credit. We solicited stimulus content for 43 items in total. These items consisted of 28 mental states which came from eight categories: four beliefs (*believe that, conclude that, feel that, think that*), four desires (*crave, desire, hope, want*), four emotions (*anger, anxiety, embarrassment, happiness*), four intentions (*goal, intend, plan, resolve*), four deliberations (*consider, deliberate, speculate, think about*), four evaluations (*value, love, hate, appreciate*), two imaginations (*imagine, visualize*), and two memory events (*forget, remember*). In addition to these 28 mental states, we included five intentional acts (*play with, eat, say, search for, avoid*), five accidents (*fall off of, trip over, slip on, run into, drop*), and five uncontrollable behaviors (*sneeze, yawn, sweat, shiver, faint*) as foils.

Subjects were provided with sentence fragments containing an ambiguous subject and a mental (or behavioral) verb, but no object (e.g., “He believed that . . .”, “She wanted . . .”, “He intended to . . .”). They were instructed to complete each sentence fragment in a way that made sense given the words provided and to avoid humor. The 28 target mental states were split across five lists and combined with the 15 observable behaviors (which were the same across all lists) and 12–13 filler trials, which included other mental phenomena such as *seeing, hearing*, and so on. This yielded approximately 33 items per list. Subjects were randomly assigned to one of these lists, which due to unbalanced randomization, yielded 13–17 contents for each item.

Unsurprisingly, many of the topics subjects wrote about were relevant to their lives as undergraduate students. Topics included concerns about school (e.g., “She felt anxious about her upcoming exam,” “He planned to do better on the next test”), romantic relationships (e.g., “She felt angry with her boyfriend,” “She thought that she wasn’t good enough for him”), and food (e.g., “He craved chocolate,” “She thought about the lunch she would be having soon”).

For the rating task, we selected five scenarios for each of the 28 mental states and 15 behavioral foils based on the most frequent contents. Any content that more than one subject provided was automatically included. For items that did not produce five pairs of duplicate contents, we selected nonduplicate contents by attempting to maximize the differences in content among the set of contents. We used this same criterion to select between scenarios when there were more than five pairs of duplicate responses for a particular item. With five items for each of the 28 mental state verbs, and for each of the

15 observable behavior foils, there were 140 mental state items and 75 observable behavior items in total. See Appendix A for a complete list.

Rating task. One hundred forty-three University of Pennsylvania students (94 female) were recruited for an experiment about “understanding others’ behavior” and completed the task for course credit. The number of subjects was set by how many students volunteered by the end of the semester. No subjects were excluded. This sample size yielded more than 90% power to detect small (*d* = .3) differences between conditions.

We quasi-randomly distributed the 215 items across five lists, such that each list contained 43 items: one of the five items from each of the 28 mental state verbs, and one of the five items from each of the 15 observable behavior foils. In a couple of cases we manually moved an item to another list to avoid the same list having two mental states with extremely similar content. Our goal with this procedure was to reduce the burden on subjects of rating many items, and to ensure that each subject rated a variety of mental state and observable behavior items, without repetition of similarly themed content. At the beginning of the experiment, subjects were randomly assigned to one of the five lists. Each item was presented on a separate page in a new random order for each subject.

Subjects responded to eight questions about each item. To minimize ambiguity, all questions contained explicit reference to the target mental (or physical) behavior (in the example below, a student believing she did well on an exam). Four questions assessed how much control subjects attributed to the agent described in each item. Two assessed the agent’ s general control: (a) “How much control did she have over *believing that she did well on the exam*?” (1: *no control at all*, 7: *complete control*; italics included in the original materials), and (b) “How much do you agree with the following statement: If she had wanted to, she could have not *believed that she did well on the exam*?” (1: *completely disagree*, 7: *completely agree*). Another two probed the agent’s intentionality: (c) “Did she intentionally *believe that she did well on the exam*?” (1: *definitely not intentionally*, 7: *definitely intentionally*), and (d) “Did she choose to *believe that she did well on the exam*?” (1: *definitely did not choose*, 7: *definitely chose*). Two additional questions probed subjects’ moral evaluations of the mental state, including (e) “How good or bad was it that she *believed that she did well on the exam?*” (-3: *very bad*, 0: *neither good nor bad*, +3: *very good*) and (f) “Should she have *believed that she did well on the exam?*” (-3: *definitely should not have*, 0: *neither should nor should not have*, 3: *definitely should have*). Two final questions probed judgments of the agent themselves: (g) “How responsible was she for *believing that she did well on the exam*?” (1: *not responsible at all*, 7: *completely responsible*), and (h) “How much does it reveal about her that she *believed that she did well on the exam*?” (1: *reveals nothing at all*, 7: *reveals a lot*). All questions used a 7-point rating scale, and were presented in a random order for each item.

At the end of the experiment, subjects reported demographic information including age, sex, political orientation, religiosity, and religious affiliation.

Results

Data Preparation. We first examined subjects’ responses to our control and intentionality measures for each of the 28 mental state items (e.g., “think that,” “believe that,” “feel angry,” etc.). Within each of the five lists, we calculated the average response across subjects for each of the 28 mental states. We then calculated correlations between our two intentionality (and, separately, two control) measures using item means. Within each of the five lists, our two intentionality measures were highly correlated with each other (*r*s within each of the five lists ranged from 0.97– 0.99, *df* = 26), as were our two control measures (*r*s = 0.93– 0.96, *df* = 26). We therefore combined them into composite measures of intentionality and control, respectively. These composite measures correlated with each other highly within each of the five lists (*r*s = .92–.96).

Within each list, we next computed subject-level averages of the composite control and intentionality ratings, ratings of the goodness or badness of the mental state (hereafter: "moral status"), and ratings that the person should or should not have this mental state (hereafter "should status"), for each of the eight mental state categories (e.g., *belief*, *desire*, etc.) and three behavior categories (*intentional, unintentional, uncontrollable*). Across the five lists, subjects' relative ratings of the 11 categories were highly correlated for both control (*alpha* = .98) and intentionality (*alpha* = .98), so we combined the five lists into a single dataset (*N* = 143, 44 means per subject: 11 item categories by four measures: control, intentionality, moral status, should status). Table 1.1 shows means and standard deviations for each category (mental states and observable behavior foils).

Table 1.1
Means (and SD) for control and intentionality composite variables, in Study 1.1.

Category	**Control**	**Intentionality**
Uncontrollable Act	2.51 (1.53)[a]	2.13 (1.40)[a]
Accidental Act	3.30 (1.51)[b]	2.29 (1.41)[b]
Memory	3.53 (1.52)[c]	2.94 (1.52)[c]
Emotion	3.71 (1.51)[d]	3.32 (1.58)[d]
Desire	4.03 (1.63)[e]	4.11 (1.71)[e]
Belief	4.54 (1.56)[f]	4.50 (1.61)[f]
Evaluation	4.59 (1.63)[f]	4.62 (1.64)[f]
Deliberation	4.99 (1.43)[g]	5.04 (1.39)[g]
Imagination	5.05 (1.35)[g]	5.16 (1.37)[g]
Intention	5.88 (1.22)[h]	5.97 (1.15)[h]
Intentional Act	5.98 (1.18)[h]	5.99 (1.13)[h]

Note: Within each column, superscripts denote means that are significantly different from each other.
Response scales ranged from 1-7.

We conducted a series of paired *t* tests on subjects' mean control and intentionality ratings between each of the eight mental state categories and the three behavioral foil categories. This allowed us to test whether subjects judged mental states as equivalently intentional or controllable to involuntary behaviors, accidental behaviors, or intentional behaviors. In this study, as well as all of the ensuing studies, we adjusted for multiple comparisons within each control measure using the Holm-Bonferroni technique. We report adjusted p values (p_{a}) which in some cases were truncated at $p_{\mathrm{a}} = 1$. For all comparisons we considered adjusted p values below .05 as statistically significant. To obtain effect size estimates, we calculated the correlated standardized mean differences (d_{rm}) using the formula recommended by Borenstein, Hedges, Higgins, and

Rothstein (2011):

$$d_{rm} = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{SD_1^2 + SD_2^2 - 2 \times r \times SD_1 \times SD_2}} \times \sqrt{2(1-r)}$$

where *r* is the correlation between pairs of observations.

Judgments of behavior foils. Subjects judged controllability and intentionality in the expected way. Uncontrollable acts were rated low on both control and intentionality whereas intentional acts were rated highly (see Table 1.1). Furthermore, accidents were rated as more controllable than uncontrollable acts, $t(142) = -9.35$, $p_a < .001$, 95% CI [-0.95, -0.62], $d_{rm} = 0.70$, but not as more intentional than uncontrollable acts, $t(142) = -2.70$, $p_a = .433$, 95% CI [-0.27, -0.04], $d_{rm} = 0.14$). There was a larger difference between judgments of the controllability of unintentional and accidental acts than between judgments of their intentionality, $F(1, 248) = 46.74$, $p = .001$. Thus, it appears that our subjects made sensible control and intentionality judgments about the behavioral foils.

Comparing mental states to behaviors. All mental states (except for intentions) were rated as less controllable and intentional than intentional acts (p_as < .001). After correcting for multiple comparisons, intentions were not significantly different than intentional acts for both control, $t(142) = -2.03$, $p_a = 1$, 95% CI [-0.20, -0.003], $d_{rm} = 0.11$, and intentionality, $t(142) = -0.49$, $p_a = 1$, 95% CI [-0.12, 0.07], $d_{rm} = 0.03$. All mental state categories were rated as more intentional than accidents (p_as = .001), and more controllable than uncontrollable behaviors (p_as = .001). Thus, in general, mental states were judged as neither fully controllable or uncontrollable, nor fully intentional or unintentional.

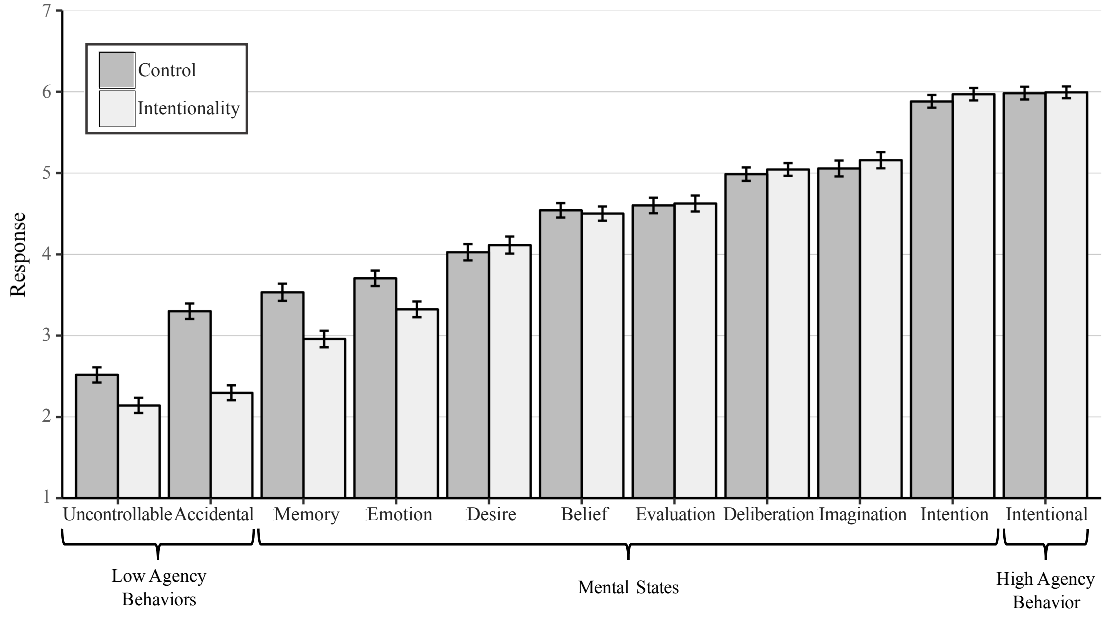

Figure 1.2. Means (and standard errors) of control and intentionality ratings in Study 1.1.

Differences between mental states. We next compared the means of each mental state category with every other mental state category. We report here the comparisons of the adjacent categories depicted in Figure 1.1. All nonadjacent categories (e.g., imaginings-beliefs) were rated significantly different from one another on both control and intentionality. Besides beliefs and evaluations, and deliberations and imaginations, all adjacent mental state categories were significantly different from one another in overall control and intentionality (see Appendix B Tables B.1 and B.2 for statistical tests).

Responsibility and character. Our secondary goal was to analyze the relationship between the control judgments and judgments of how responsible the agent was for the mental state, and how much the mental state revealed information about who the person is (character-relevance). We first analyzed the relationship between these three variables across subjects' average ratings for all eight mental state categories (eight observations per subject, per judgment type). We regressed responsibility and character-relevance judgments on judgments of control (and separately, intentionality) using linear mixed-effect models. The final models included control (or intentionality), moral status, should status, as well as random intercepts for subject, and random slopes for control (or intentionality), moral status, and should status. We excluded data from the uncontrollable, accidental, and intentional behavior categories, since our focus here was solely on judgments about mental states. We found that both kinds of control strongly predicted responsibility judgments (control: $b = 0.71$, $SE = 0.03$, $t = 28.89$, $p < .001$; intentionality: $b = 0.62$, $SE = 0.02$, $t = 26.18$, $p < .001$), as well as character-relevance judgments (control: $b = 0.19$, $SE = 0.03$, $t = 6.82$, $p < .001$; intentionality: $b = 0.17$, $SE = 0.03$, $t = 6.33$, $p < .001$). We observed the same pattern of results when we conduct these analyses using subjects' average ratings to the 28 specific mental state items rather than the eight superordinate categories.

Although intentionality and control predicted both responsibility and character judgments, the relationship appeared considerably stronger for responsibility. To test whether the strengths of these relationships differed, we created a single new variable, "social judgment," which contained separate responsibility and character judgments for each subject—each subject contributed eight responsibility judgments (for each of the eight mental states) and eight character judgments to this variable. We also created another binary variable, "attribution type," which coded whether the judgment was of responsibility or character-relevance. In two separate analyses, we then regressed the social judgment variable on control (or intentionality), attribution type, and the interaction of control (or intentionality) and attribution type. The final model also included by-subject (and by-mental state category) random slopes as well as random intercepts for the attribution type by control interaction. These analyses revealed interactions between control and attribution type ($b = 0.461$, $SE = 0.039$, $t = 11.806$, $p < .001$), and between intentionality and attribution type ($b = 0.383$, $SE = 0.038$, $t = 10.09$, $p < .001$), thereby showing that control and intentionality judgments were indeed more strongly correlated with responsibility judgments than with character judgments.

We next looked specifically at emotions, desires, beliefs, and evaluations, as these mental states are often treated as diagnostic of character and are featured prominently in arguments that that control is unnecessary for moral responsibility. As can be seen in Figure 1.3, the positive relationships between control and responsibility, and control and character, replicates within each of these four categories (statistical tests reported in Table 1.2). Additionally, we found within each of these four mental states that responsibility was more strongly correlated with control than character was with control (see Table 1.2).

Table 1.2
Correlations between agency (control or intentionality), responsibility and character in Study 1.1, as well as test for difference in size of correlation between agency and responsibility (A) and agency and character (B).

Mental State	**Agency Type**	**A: Agency x Responsibility**	**B: Agency x Character**	**Responsibility x Character**	**A vs B (t-value)**
Emotion	Control	0.77	0.27	0.36	7.70
	Intentionality	0.73	0.23	0.36	7.37
Desire	Control	0.84	0.34	0.35	8.79
	Intentionality	0.83	0.33	0.35	8.60
Belief	Control	0.82	0.36	0.36	7.83
	Intentionality	0.78	0.34	0.36	6.96
Evaluation	Control	0.78	0.52	0.52	4.91
	Intentionality	0.82	0.55	0.52	5.56

Notes. *df* = 141. All correlations and *t-values* significant at $p < 0.001$.

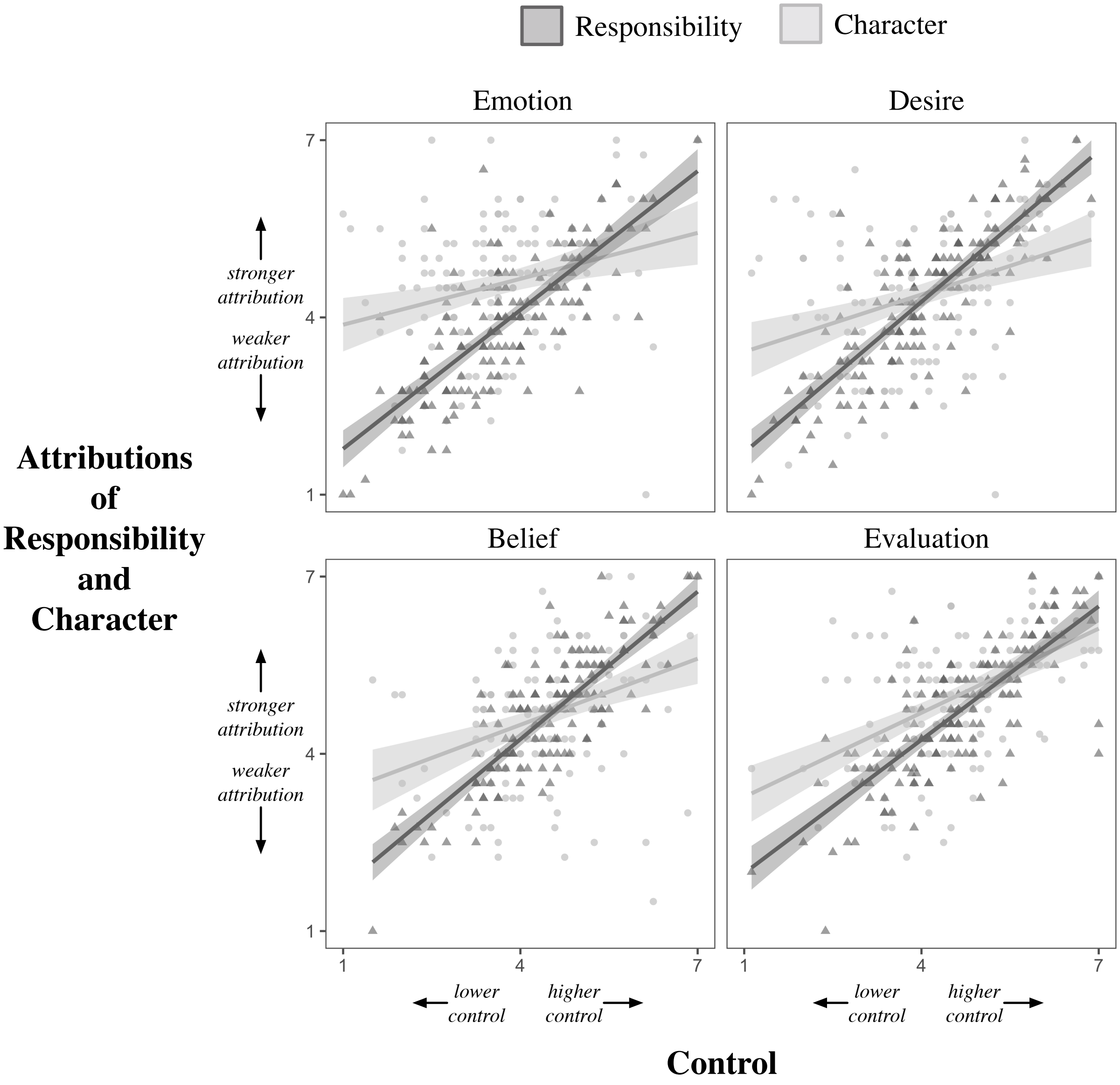

Figure 1.3: Individual differences in average attribution of control across emotions, desires, beliefs, and evaluations correlate with attributions of responsibility (triangles/darker) and character relevance (circle/lighter).

Discussion

The purpose of Study 1.1 was to discover whether a range of representative mental states are typically judged as intentional and controllable, or not. Our results revealed that people attributed surprisingly high levels of control over mental states—moderate-to-high degrees of control and intentionality for many mental state categories, including desires, beliefs, and evaluations. Even emotions were judged to be somewhat controllable and intentional. Thus, these results conflict with the claim that many mental states are judged as unintentional (cf. Malle & Knobe, 1997b), or as merely "happening" (cf. Gilovich & Regan, 1986). We did, however, replicate prior studies showing that mental states are rated as less intentional and chosen than intentional behaviors (Gilovich & Regan, 1986; Malle & Knobe, 1997a), consistent with views summarized above that emotions, desires, and evaluations are not seen as willfully controllable (e.g., Adams, 1985; D'Andrade, 1987).

Study 1.1 also revealed reliable differences across mental state categories in the amount of control people attributed, supporting the idea that people treat different types of mental states as varying in control and intentionality (D'Andrade, 1987). For instance, deliberations and imaginations were judged as highly intentional and controllable— more controllable than beliefs, desires, emotions, evaluations, and memories. Although we observed differences in control across mental states, the content of the mental states was not held constant. That is, desires tended to be about different things than beliefs, evaluations, and so on. Differences in rated control may have resulted from differences in the content of each mental state, rather than from more fundamental conceptual differences between mental state categories. We address this in Study 1.2.

Finally, Study 1.1 provided preliminary evidence that attributions of responsibility are associated with attributions of control. Subjects who thought that emotions, beliefs, desires, and other mental states were uncontrollable tended to also say that the target was not responsible for the mental state (see, e.g., Figure 1.3). Though control (and intentionality) also correlated with character judgments, they did so to a much lesser degree. Investigating Figure 1.3, subjects who judged mental states as highly uncontrollable attributed practically no responsibility to the target over the state, while still judging that the mental state was somewhat diagnostic of that person's character. This is exactly what would be predicted by control theories of moral responsibility: even though mental states are diagnostic of character, they are not thereby automatically things that the mental state holder is responsible for. Indeed, in a set of exploratory analyses, control (and intentionality) more strongly correlated with responsibility than character diagnosticity did, suggesting that control is a more important input to responsibility judgments than character is. None of these findings would be predicted by a character account of responsibility. However, in Study 1.1, subjects made judgments about subject-generated mental states, most of which were neutral or non-moral in nature. In Study 1.2 we investigated how people others' control over, and moral responsibility for, immoral mental states.

Study 1.2

In Study 1.2 we tested whether the association between control and moral responsibility we observed in Study 1.1 replicates when people evaluate others' *immoral* mental states. In these contexts, as opposed to the more neutral stimuli we tested in Study 1.1, the mental state may be more likely to be seen as highly diagnostic of someone's negative immoral character. Thus, it could be in these contexts that control plays little important role in subjects' evaluations of moral responsibility. Thus, we tested whether control strongly predict attributions of moral responsibility when people evaluate others' immoral mental states.

Study 1.2 differed from Study 1.1 in one other important way. Rather than rely on subject-generated mental states, we wrote four scenarios in which a target holds an immoral emotion, desire, belief or evaluative attitude. Although this method sacrifices some degree of external validity, it allows us to make two important changes. First, we can specify the context of the situation in greater detail than what subjects were provided in Study. It is possible that greater situational context decreases perceived control because people are more aware of the environmental causes of the mental state. However, it is also possible that people perceive a greater degree of intentionality and control, because alternative reactions seem open to him or her. Either way, providing greater background detail about the target and the mental state better mimics the state of affairs in which people evaluate immoral mental states in real life; namely, with some awareness of the situation in which they occur. Second, in the vignettes we specified that the mental state that the target forms is in relation to something over which he or she has little control. This design choice is similar to that used by others to ensure that subjects' reactions are not based on their fear that the target will have some influence over whether harm occurs (c.f. Inbar et al., 2012).

We also adopted the secondary objective of testing whether the differences in mental state control we observed in Study 1.1 replicated. One possible explanation for why people judged emotions and desires to be less controllable than beliefs and evaluative states is because these mental states are seen as inherently less controllable. However, it could just be that the types of prototypical emotions and desires that subjects generated in Study 1.1 were associated with less controllable content, or less controllable contexts, than the beliefs and evaluative attitudes that they reasoned about. To address this, we created multiple versions of each vignette which held the content the mental state the same but varied what mental state type subjects judged.

Participants. Two hundred subjects (98 female, M_{age} = 34.3) were recruited from Amazon's Mechanical Turk. No subjects were excluded from our analyses. To mask the purpose of the experiment, subjects were told that the study was about understanding behavior and that they would read four stories about a person before making a series of judgments about that person. Sample size was determined prior to data collection, using the same simulation procedure we used for Study 1.2, adjusted in light of changes to the experimental design. This analysis revealed that a sample size of 200 would yield > 90% power to detect absolute mean differences comparable to those in prior studies (b = .30).

Stimuli. We constructed four vignettes in which a target character learns about a state of affairs over which they have little or no control and has a nonnormative response. This response was either a negative emotional reaction (feeling "upset" or "angry"), a desire ("wanting" or "desiring" a different state), a belief ("thinking" or "believing" something), or an evaluative attitude ("disliking" or "hating" something about the state of affairs). See below for the full text of one scenario, with each of the possible nonnormative responses listed (subjects judged only one such response per vignette).

> James is a 50-year-old White male. He grew up in a middle-class family and is currently a manager at a bank. He married a few years after graduating college and he and his wife have a daughter. James's daughter is currently living and

> working in another state and has just called to tell her parents she has entered into a serious relationship. Over the course of the phone call it becomes clear that her boyfriend is African American. When he hears this, James . . .
>
> > feels unhappy/angry that his daughter is dating an African American.
> >
> > desires that/wants his daughter not be/not to be dating an African American.
> >
> > believes/thinks that it is wrong for his daughter to be dating an African American.
> >
> > hates/dislikes that his daughter is dating an African American.

The three other scenarios (with the antisocial desires presented as illustrative) involved a man learning that his mother was involved in a car accident and not wanting her to survive, a civilian learning about a UN military operation designed to block murderous terrorists and not wanting this mission to succeed, and a student watching video footage of a journalist being tortured and wanting the journalist to be in more pain (for full details, see Appendix C).

Assignment of conditions. The four mental state types (emotion, desire, belief, evaluation) were crossed with the four scenarios in a Latin Square design. This resulted in four lists. Each list comprised four unique pairings of the four mental states and scenarios. Within each list, there was one instance of each mental state, and one instance of each scenario, and there was no repetition of any mental state-scenario pairing across the four lists. At the beginning of the survey, subjects were randomly assigned to one of the four lists. The order of presentation for each scenario was randomly determined for each subject.

Subjects were also randomly assigned to see one of two possible mental states for each item within each list (e.g., "unhappy" or "angry" in the emotion condition); however, randomization was weighted such that, at the end of the experiment, both mental states within each mental state condition were shown to an equal number of subjects.

Procedure. Subjects answered five questions for each of the four scenarios. For clarity, each question presented the content of the mental state in italics and the relevant control construct in bold. Subjects judged the intentionality of the mental state (e.g., "Did James **intentionally** *feel angry that his daughter is dating an African American*?" 1: *definitely not intentionally*, 7: *definitely intentionally*), whether the agent could choose to stop having the mental state ("Can James **choose to stop** *feeling angry that his daughter is dating an African American*?" 1: *definitely cannot choose to stop*, 7: *definitely can choose to stop*), the wrongness of the mental state ("How **morally wrong** is it for James to *feel angry that his daughter is dating an African American*?" 1: *not morally wrong at all*, 7: *extremely morally wrong*), the agent's blameworthiness ("How **blameworthy** is James for *feeling angry that his daughter is dating an African American*?" 1: *not blameworthy at all*, 7: *extremely blameworthy*) and the agent's character ("How bad is James's **moral character** for *feeling angry that his daughter is dating an African American*?" 1: *not bad at all*, 7: *extremely bad*). All ratings were made on a seven-point rating scale ranging from 1 to 7. The order of the questions was randomly determined for each trial.

Results

See Table 1.3 for means and standard deviations for each of the dependent measures.

Table 1.3
Means (and standard deviations) for dependent variables in Study 1.2 by mental state condition.

Mental State	Intentionality	Stop	Blameworthiness	Character	Wrongness
Emotion	4.69 (1.89)	5.26 (1.61)	5.30 (1.71)	5.47 (1.54)	5.58 (1.54)
Desire	5.62 (1.66)	5.49 (1.68)	5.46 (1.79)	5.66 (1.57)	5.71 (1.57)
Belief	5.15 (1.77)	5.49 (1.70)	5.20 (1.77)	5.38 (1.58)	5.48 (1.62)
Evaluation	5.17 (1.80)	5.39 (1.70)	5.24 (1.76)	5.47 (1.59)	5.62 (1.65)

Note: Ratings were made on a seven-point scale anchored at 1 and 7.

Blameworthiness and character. We conducted a series of four linear mixed-effect models (LMEM) to investigate whether judgments of intentionality (or, separately, the ability to stop) predicted attributions of blameworthiness and character. Each model included a fixed effect of our control DV (either intentionality or stop) as well as judgments of wrongness. We included wrongness as a predictor because subjects may have differed in how morally objectionable they rated each mental state to be, which would then also impact how blameworthy the person is for having the mental state, and how negative the target's moral character was (c.f. Cushman, 2008). In addition, each model included subject and item intercepts for the attribution judgment (either blameworthiness or character). Unsurprisingly, wrongness was a significant predictor of both blameworthiness and character (*p*s < .001). Intentionality was also a significant predictor of both blameworthiness ($b = 0.35$, $SE = 0.03$, $t = 13.11$, $p < .001$) and character ($b = 0.15$, $SE = 0.02$, $t = 8.90$, $p < .001$), as was the ability to stop the mental state (blameworthiness: $b = 0.44$, $SE = 0.04$, $t = 12.09$, $p < .001$; character: $b = 0.33$, $SE = 0.03$, $t = 9.74$, $p < .001$). These results replicated our finding from Study 1.1 showing that attributions of control predict moral judgments of blameworthiness and character even in the domain of highly immoral mental states.

We then repeated the analysis described in Study 1.1 testing whether control correlated more strongly with blameworthiness than with character. We aggregated blameworthiness and character judgments into a new dependent variable, attribution response, predicted by a new independent variable, attribution type. In our first analysis we regressed attribution response on wrongness, intentionality, attribution type, and the interaction of intentionality and attribution type. Similar to the LMEM above, models contained by-subject and by-vignette random intercepts as well as by by-subject and by-vignette random slopes for each of the predictors except for perceived wrongness (which was removed to avoid singular fit). This analysis returned a significant interaction, b = -0.11, SE = 0.04, t = -2.95, p < 0.001, revealing that intentionality ratings more strongly predicted blameworthiness than character. The same analysis conducted on stop ratings also revealed a significant interaction, b = -0.10, SE = 0.04, t = -2.55, p = 0.01, indicating that ratings of the ability to stop having the mental state more strongly predicted blameworthiness judgments than did character judgments. Identical analyses without including wrongness as a predictor yield the same results.

However, when we conducted the same analysis looking within each mental state, we only observed a reliable dissociation between blameworthiness and character for emotions, see Table 1.4. One explanation for this is that, for non-emotion mental states, intentionality, stop, blame, and character ratings were all very high, restricting the degree of variation and reducing power; see Table 1.4.

Table 1.4
Correlations between agency (control or intentionality), responsibility and character in Study 1.1, as well as t-value for difference in size of correlation between agency and responsibility (A) and agency and character (B).

Mental State	Agency Type	A: Agency x Blameworthy	B: Agency x Character	Blameworthy x Character	A vs B (t-value)
Emotion	Stop	0.48**	0.31**	0.66**	3.26*
	Intentionality	0.62**	0.41**	0.66**	4.49**
Desire	Stop	0.42**	0.40**	0.70**	0.40
	Intentionality	0.56**	0.50**	0.70**	1.32
Belief	Stop	0.44**	0.38**	0.68**	1.17
	Intentionality	0.51**	0.36**	0.68**	3.03*
Evaluation	Stop	0.33**	0.26**	0.67**	1.27
	Intentionality	0.51**	0.43**	0.67**	1.63

Notes. ** $p < 0.001$. * $p < 0.01$

Intentionality and stop analyses. Our design also allowed us to test whether the differences in control across emotions, desires, beliefs, and evaluations that we observed in Study 1.1 replicated once the context and content of the mental state was held constant. To test this, we conducted separate LMEM regressions for the intentionality and choose to stop variables, including every trial (four per subject), random intercepts for subject and scenario, and random slopes for mental state by scenario. In each model we regressed intentionality (or stop) on a single predictor, our four-category mental state variable, with a priori contrasts between emotion and desire, desire and belief, and belief and evaluation. Replicating Study 1.1, emotions were rated as less intentional than desires ($b = 0.92$, $SE = 0.13$, $t = 7.02$, $p < .001$). However, in contrast to Study 1.1, desires were rated as *more* intentional than beliefs ($b = -0.46$, $SE = 0.17$, $t = -2.70$, $p = .01$). Beliefs and evaluations did not differ ($b = 0.03$, $SE = 0.15$, $t = 0.21$, $p = .83$). Ratings of whether the agent could choose to stop having the attitude were less differentiated: there was no significant difference between emotions and desires ($b = 0.21$, $SE = 0.18$, $t = 1.19$, $p = .235$), and no difference between desires and beliefs ($b = 0.01$, $SE = 0.15$, $t = 0.08$, $p =$

.937). Likewise, beliefs and evaluations were rated as similarly stoppable (*b* = -0.13, *SE* = 0.25, *t* = -0.55, *p* = .584).

Discussion

Study 1.2 replicated key findings from Study 1.1. First, subjects attributed a great deal of control to the immoral mental state holders over their emotions, desires, beliefs, and evaluations. This replication is noteworthy in light of two changes from Study 1.1. First, the mental states were now highly immoral, rather than neutral or non-moral. It was possible, for instance, that upon reading about someone who likes seeing someone in pain that people interpret that attitude as pathological, or the sign of a damaged psyche, and therefore not controllable. Second, in these studies the target had a more information about the context in which the mental state occurs, including basic knowledge about the person with the mental state and the circumstance in which the mental state forms. It was possible that, with these details, subjects would judge the mental state as being caused by the environment, or other situational forces, rather than by the person. However, despite these changes, subjects by-and-large judged that the target intentionally chose to experience the emotion, belief, desire, or evaluative attitude that they did, and that they could choose to stop having or feeling it if they wanted.

Study 1.2 also replicated an apparent dissociation between judgments of moral responsibility, in this case blameworthiness, and judgments of character diagnosticity. As in Study 1.1, ratings of control, either intentionality or the ability to stop having the mental state, more strongly predicted blameworthiness judgments than they did character judgments. One notable difference between Study 1.2 and Study 1.1, however, is that blameworthiness and character were much more strongly correlated with one another. This is unsurprising. Both are strongly predicted by how immoral the mental state is such that the more egregious one's mental state the more blame one deserves for it, as well as the more diagnostic it is of poor moral character. Perhaps for this reason, as well as the high amount of control, blame, and poor character subjects attributed to the target, the dissociation between blameworthiness and character was not reliably found when we looked within mental states. We return to this in Study 1.3.

Finally, our design allowed us to test whether the differences in mental state control that we observed in Study 1.1 replicated once the context and content where closely matched. To our surprise, we found that they largely did not. When looking at the ability to stop holding a mental state, we found no difference between any of the mental states: subjects judged that the targets in the vignettes were all highly capable of no longer having the mental state if they wanted. It was only when rating intentionality that we observed any reliable differences between mental states. Similar to Study 1.1, subjects judged emotions as less intentional than desires, beliefs, and evaluations. Interestingly, in this study, desires were rated as more intentional than beliefs and evaluations. Overall, it appears that the differences we observed in Study 1.1 reflect the prototypical emotions, desires, and beliefs that people spontaneously think about, rather than reflect differences intrinsic to lay people's conception of emotions, desires, and beliefs.

The data presented so far suggest that everyday mental state responsibility poses little challenge to the control theory of responsibility. First, people by-and-large attribute intentional control to others over their mental states. Especially compared to intuitively uncontrollable behaviors like blushing or sneezing, people believe others deliberately choose to adopt the emotions, desires, beliefs, and evaluative attitudes that they do; and furthermore, believe that others can intentionally change their attitudes if they want to. Thus, the apparent observation that people regularly hold each other responsible for things outside of their control appears to be unfounded. Second, attributions of control strongly correlate with responsibility (Study 1.1) and blame (Study 1.2).

We build on these results in Study 1.3. We reasoned that if moral responsibility required attributions of control, and that if character diagnosticity did not (or did to a lesser degree), then it should be possible to change subjects' attributions of responsibility by manipulating control without *also* changing the perceived character diagnosticity of the mental state. Doing so would not only provide causal evidence for the role of control judgments of moral responsibility for mental states but provide additional evidence against models of moral evaluation in which character diagnosticity is deemed necessary or sufficient.

Study 1.3 builds on our results in one other important way. Control theories of moral responsibility predict not only that certain kinds of moral judgments (like blameworthiness) require attributions of control, but that certain behavioral reactions, like punishing, confronting, or making demands, require that the target had control over the transgression in question. In Study 1.3 we expanded our list of dependent measures to include a series of behavioral reactions we hypothesized would be strongly predicted by control, like confronting someone over their mental state, and some that we predicted that would be weakly predicted by control, like avoiding someone who has an immoral mental state. By extension, we further hypothesized that manipulating control would affect people's reported likelihood of performing behaviors that constitute holding someone responsible but not behaviors which did not involve holding someone responsible.

Study 1.3

Our primary goal in Study 1.3 was to manipulate perceived control over the mental state and test the effect that this has on downstream reactions to an immoral mental state. To manipulate perceived control, we took advantage of the fact that ordinary people believe that certain types of causal explanations for someone's behavior, like someone's genetics, biology, or tragic life histories, entail lower control over the behavior in question. For instance, past research has shown that bad behavior (e.g., gang membership), physical illness (e.g., obesity), or mental illness (e.g., schizophrenia) that is caused by someone's genes, brain chemistry, or tragic past, is judged as less controllable, and the target is judged less blameworthy for it (see, e.g., Cheung & Heine, 2015; Dar-Nimrod, Heine, Cheung, & Schaller, 2011; Gill & Cerce, 2017; Lebowitz & Ahn, 2014; Lebowitz, Rosenthal, & Ahn, 2016; Monterosso, Royzman, & Schwartz, 2005). We predicted that similar explanations would reduce perceived control over an immoral mental state as well. We further predicted that, consistent with the findings from Studies 1 and 2, reduced control would result in reduced blameworthiness as well as a reduced comfort confronting the individual for their mental state.

And finally, we predicted that reducing control would have a smaller, possibly negligible, effect on attributions of poor moral character and a desire to avoid the target. This latter prediction is critical for adjudicating between the control and character models of moral judgment. After all, if every change in perceived control and blame is associated with a change in perceived poor character, then it is possible that character attribution is the root cause of people's moral evaluation, and that blameworthiness and control are by-products of this attribution. However, if it is possible to reduce attributions of moral responsibility by a reduction in perceived control, without changing perceived character, then it would show that poor character evaluation is not sufficient to license moral responsibility.

Study 1.3 used similar immoral mental state vignettes as those used in Study 1.2. Because Study 1.2 showed very few differences across mental state types, we changed the text of immoral mental state to always be described as an evaluative mental state – specifically, liking or disliking. In Studies 1 and 2, evaluative mental states tended to be judged as highly intentional and controllable and, in Study 1.2, as similar to desires and beliefs. Additionally, in Study 1.2 we failed to observe a reliable dissociation between blameworthiness and character for evaluations, desires, or beliefs. Thus, this study is an especially conservative test of our prediction.

Methods

Participants. We recruited 269 college undergraduates (140 reported female) from a university on the East Coast who were compensated with course credit. This sample size reflects the total number of subjects who volunteered for the study before the end of the semester.

Design. Our study used a 2 condition (constraining causal history vs non-constraining causal history) design which we replicated across four vignettes. We constructed two lists which each contained two vignettes with the constraining causal history, and two vignettes with the unconstraining causal history. Subjects were randomly assigned to one of the two lists at the beginning of the study.

Materials and Procedure. Subjects read and reacted to four vignettes that described someone who had an immoral mental state. For instance, one vignette describes a college aged kid named Paul who is struggling with his grades and whose mother has been on his case to study more. The mother decides to visit her son but gets in a car wreck. When Paul learns about the wreck and told that the doctors are uncertain about whether or not she will live, he likes the idea that she might die. Other vignettes featured mental states like disliking that one's daughter is engaged to an African American, disliking that a rescue mission on TV would succeed, and liking that a journalist had been badly tortured. See Table 1.5 below for the exact text of each immoral mental state.

Table 1.5
Description of immoral mental states, text for “constraining” condition and corresponding text in the “non-constraining” condition.

Mental State	**Constraining Causal History**	**Non-constraining Causal History**
Paul is still thinking about what a pain his mother has been lately and, in that moment, likes the idea of his mother passing away.	Paul has a developmental disorder and, as a result of this disorder, lacks the ability to feel empathy for others or form normal familial bonds with them.	Paul has a developmental disorder and, as a result, has difficulty speaking in fluid sentences.
Although James does not say anything to his daughter, he dislikes that his daughter is dating an African American.	James’s father was a hateful person who constantly told his children that black people were dangerous and irresponsible. All of James’s siblings have attitudes like this deeply ingrained in them.	James’s father was an overbearing busy-body who tried controlling every aspect of his children’s lives. All of James’s siblings make judgments about their children’s life choices as well.
Wesley dislikes that the UN counter attack will likely succeed.	Wesley slipped in the shower and hit his head about a year ago. While he is completely healthy again, his worldview has changed in a lot of ways. The doctors suspect that this is because his brain chemistry is different, which is affecting, among other things, his thoughts and beliefs.	Wesley slipped in the shower and broke his arm about a year ago. While he is healthy again, his movement in his arm is still restricted and is occasionally sore. The doctors suspect that his muscles will never fully recover.
Digging through some archives, she found video footage of a journalist being beaten and tortured by secret police. While watching the footage, Amy likes that the journalist is in a great deal of pain as this makes for a better senior thesis.	Her father has been pressuring her to succeed ever since she was a child to the point where her entire identity has become about school. So, matter what she does, when she thinks about the journalist’s pain, she is numb to it. Instead, her mind turns to the thought of failing her thesis, not graduating with honors, and disappointing her father.	Her father has been pressuring her to succeed ever since she was a child to the point where her entire identity has become about school. In addition to working on a senior thesis, her father has insisted that she take graduate-level coursework and run for student government.

Embedded in the vignette was background information about the person who experienced the negative mental state. In the constraining causal history conditions, subjects were given some information to suggest that the target’s immoral mental state was outside of his or her control. The specific reason varied across vignettes. In one vignette, it was because the target had a developmental disorder that prevents him from feeling empathy or forming normal familial bonds. In another vignette, the individual had suffered some brain trauma that affected his or her attitudes. Each constraining condition was paired with a non-constraining causal history condition. In this condition, the target was described as having a similar background, but not one that affects their mental state. Keeping with the example above, as opposed to a developmental disorder that affects the target's capacity for empathy, he has a developmental disorder that affects his ability to speak in fluid sentences. This was done to avoid a potential confound; namely, that subjects would reduce the severity of their moral judgments upon being given any information suggesting that the target was unlucky, victimized, or otherwise sympathetic.

Subjects reported their agreement with five statements about each vignette. One statement measured perceived control over the mental state "Paul could choose to stop having this attitude if he really wanted to". Two items measured the two moral judgments of interest, including perceived blameworthiness for the mental state "Paul is blameworthy for liking the idea of his mother passing away" and perceived character "Paul is a person with low moral character because he likes the idea of his mother passing away". And the final two questions measured behavioral intentions. One measured people's intent to confront Paul "If I knew Paul, I would confront him and try to make him feel bad for this attitude" and the other measured people's judgments that others should not form relationships with the target "People should not form close relationships with Paul". Each statement changed the name of the target, and the description of the immoral mental state, to match the text of the vignette. Subjects rated their agreement with each statement on a seven-point rating scale (1: completely disagree, 7: completely agree). Each of the dependent measures was shown in a random order. And each of the four vignettes within each list were shown in a random order for each subject.

Subjects then completed a task unrelated to this study and filled out a demographics form.

Results

Analysis Procedure. In the analysis below, we used a mixed-effect linear model with a “maximal” random effect structure (Barr et al, 2016). This meant that every model included random by-subject and by-vignette intercepts for the DV, as well as random by-subject and by-vignette slopes for condition. This means that our regression accounts for variation in mean responses to the DV, as well as variation in the efficacy of the manipulation on each DV, across subjects and across vignettes. Table 1.6, below, reports the means, standard deviations, and correlations of the five dependent measures.

Table 1.6
Descriptive statistics for Study 1.3.

	Means (and SD)		**Pearson Correlation**			
	Non-constraining	Constraining				
Measure	Causal History	Causal History	**1**	**2**	**3**	**4**
1. Control	4.84 (1.27)	3.99 (1.49)*				
2. Blameworthy	4.58 (1.30)	3.80 (1.33)**	0.54**			
3. Character	4.29 (1.24)	3.94 (1.31)	0.39**	0.58**		
4. Confront	4.24 (1.43)	3.92 (1.41)**	0.38**	0.53**	0.55**	
5. Avoid	3.27 (1.37)	3.26 (1.38)	0.11*	0.34**	0.61**	0.40**

Notes. All ratings made on a 7-point rating scale (1-7).
Values used to compute mean and SD, as well as used in correlation analyses, were obtained by computing each subject's average rating for each DV in the low and high control conditions
df = 536. * $p < 0.05$, ** $p < 0.001$

Main Analyses. We first examined overall mean differences between the constraining causal history and non-constraining causal history conditions (see Figure 1.4, below). As expected, subjects attributed less control to the target over the mental state in the constraining condition relative to the non-constraining condition, b = -0.85, *SE* = 0.31, t = -2.78, p = 0.005. Additionally, and as predicted, targets in the constraining condition were judged as less blameworthy, b = -0.78, *SE* = 0.19, t = -4.05, $p < 0.001$. Subjects were also less likely to confront the target for his or her attitude, b = -0.33, *SE* = 0.09, t = -3.80, $p < 0.001$. However, and as expected, the existence of constraints on the targets' mental states had a negligible impact on attributions of poor moral character , b = -0.35, *SE* = 0.20, t = -1.79, p = 0.074. Similarly, we observed no reliable difference in avoidance ratings across conditions, b = -0.01, *SE* = 0.26, t = -0.05, p = 0.96.

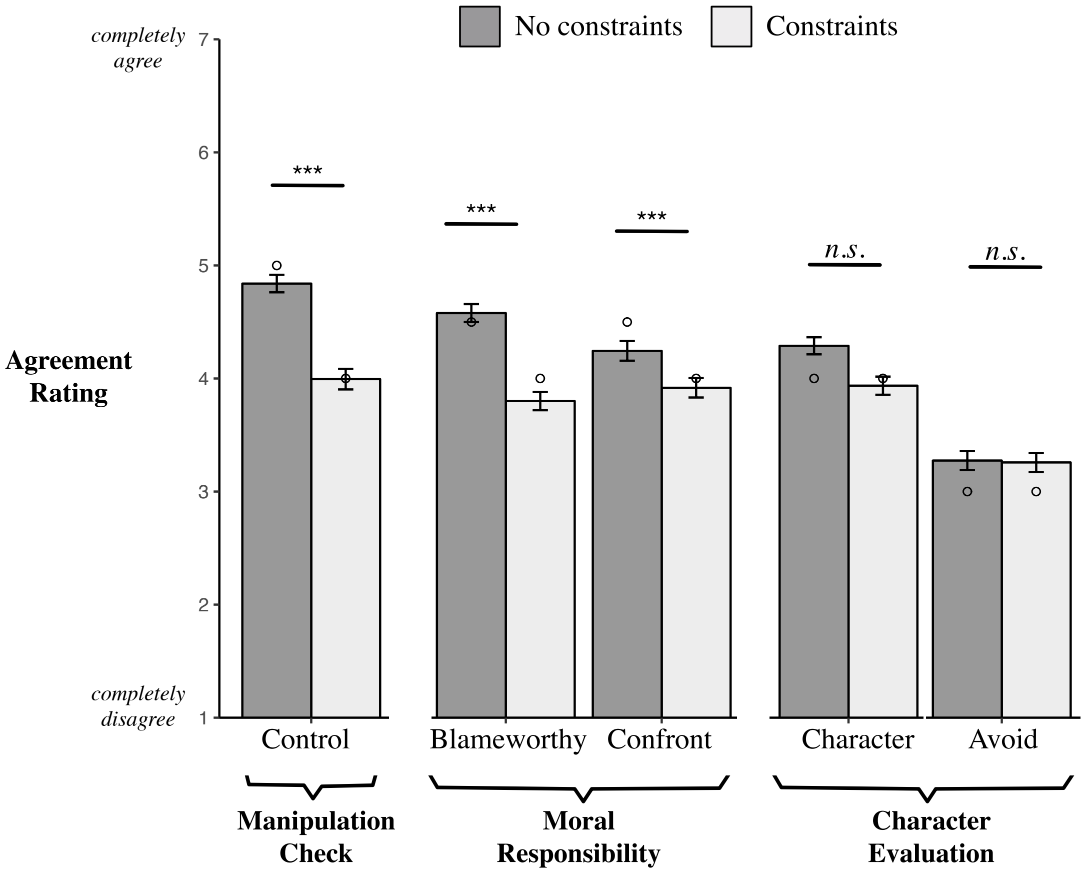

Figure 1.4: A. Means and standard error for Study 4 for dependent measures grouped by measure type. Circles represent median values. *** $p < 0.001$. *n.s.* $p > 0.05$.

We next tested whether perceived mental state control mediated the differences between the existence of constraints (or not) and subjects' reactions. To do this, we calculate subjects' average ratings for each of the five measures across the vignettes within condition. We then used these values to conduct a series of within-subjects mediation analysis (Figure 1.5). These analyses, shown in Figure 1.5a-5d, reveal that perceived mental state control mediates the observed differences in blameworthiness, confrontation, and character, but not avoidance, across condition. Thus, our findings are consistent with our hypothesis; namely, that control has a causal impact on people's moral judgments.

Our key test was whether control affected blameworthiness and confrontation more than it affected character judgments and avoidance reactions. We tested this using a technique similar to the one used in Studies 1 and 2. We aggregated subjects' blameworthiness and character ratings into a single variable Moral Judgment predicted by the independent variable Judgment Type. We then regressed moral judgment ratings on judgment type, condition, and the interaction of judgment type and condition. As above, our regression model used a maximal random effect structure. Consistent with our predictions, we observed a significant interaction of judgment type and condition, $b = 0.43$, $SE = 0.12$, $t = 3.60$, $p < 0.001$, indicating that the effect of our constraint manipulation was stronger on blameworthiness judgments compared to character judgments. We repeated this analysis on subjects' confront and avoid reactions creating a new dependent measure Behavior Rating and new independent variable Behavior Type. However, here we did not observe that the effect of the manipulation reliably depended on which behavior subjects were judging, $b = 0.32$, $SE = 0.18$, $t = 1.72$, $p = 0.085$.

The findings above suggest that the existence of causally constraining explanations for immoral mental states affects control, and affects certain kinds of reactions, like blameworthiness, more than others, like character. This suggests that it is specifically the changes in perceived control that are driving a wedge between these reactions. To test this directly, we used subjects' average responses, across vignettes but within condition and then calculated blameworthiness-minus-character difference scores, and confront-minus-avoid difference scores[1]. We predicted that the change in difference scores across condition would be mediated by changes in perceived control. That is, the greater decrease in blameworthiness, relative to character evaluation, from the no-constraint to constraint condition is correlated with the change in perceived mental state control.

To test whether these differences were predicted by the change in subjects' control ratings between conditions, we created two within-subjects mediation models, one predicting blame-minus-character difference scores (Figure 1.5e) and one predicting confront-minus-avoid difference scores (Figure 1.5f). Results showed that changes in perceived control mediated the effect of condition on blame-minus-character difference scores (a: $b = 0.85$, $p < 0.001$; b: $b = 0.22$, $p < 0.001$; c: $b = 0.44$, $p < 0.001$; c': $b = 0.24$, $p = .012$). Results from our second model showed that control mediated the effect of condition on confront-minus-avoid differences scores (a: $b = 0.85$, $p < 0.001$; b: $b = 0.39$, $p < 0.001$; c: $b = 0.31$, $p = 0.004$; c': $b = -0.02$, $p = 0.843$). This results directly support our hypothesis: namely, that, even in the context of mental state evaluation, moral responsibility reactions are distinguished from character evaluations, and are distinguished by a reliance on considerations of control.

[1] We replicated this analysis (as well as the analysis reported below) using within-subject z-scores to calculate difference scores. This is to account for the fact that subjects many have used the response scales differently across measures. Results from these analyses were identical.

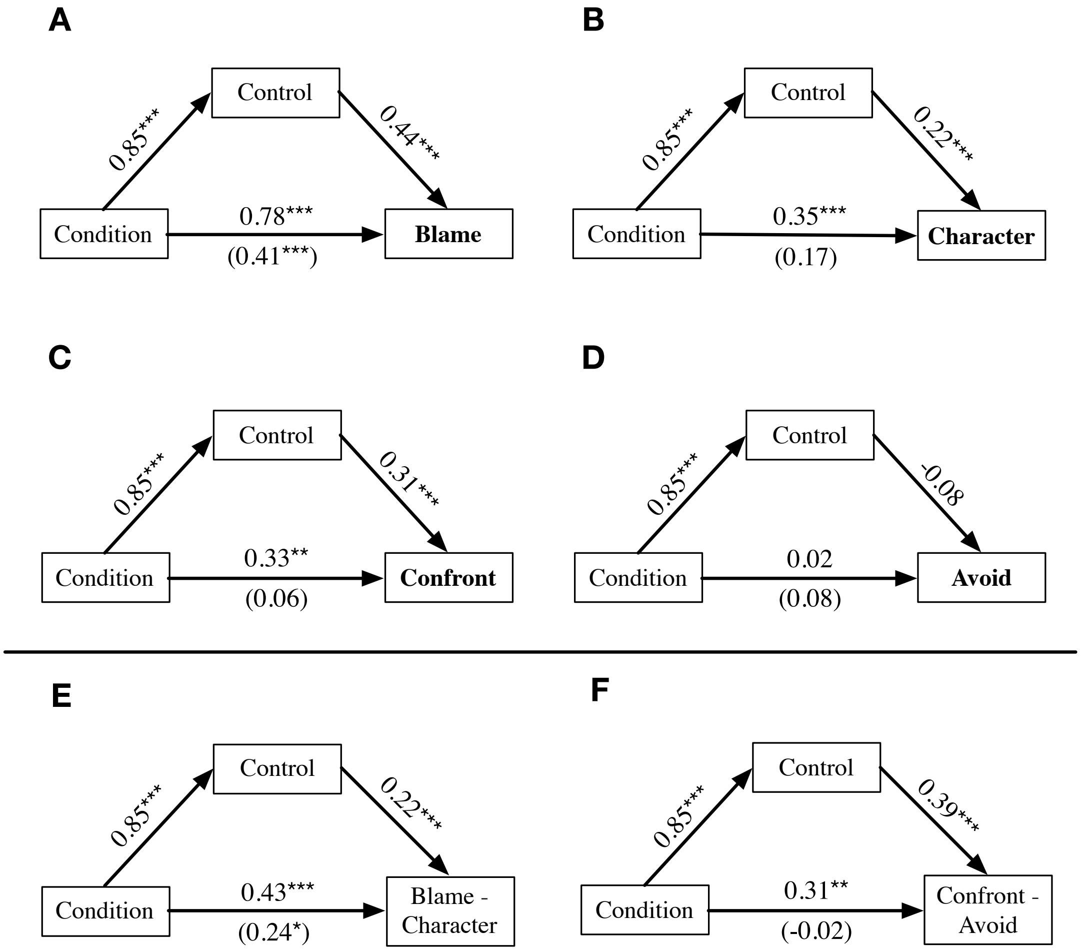

Figure 1.5. Mediation analyses examining the mediating role of perceived control on dependent measures. (A) – (D) shows control mediating the effect of constraint on blame (A) and character (B) judgments, and confront (C) and avoiding (D) intentions. (E)-(F) show that the differential effect of constraint on blameworthiness and character (E), and confrontation and avoidance (F) is mediating by perceived control. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Discussion

In Study 1.3 we successfully manipulated perceived mental state control by stipulating that the person's immoral mental state was caused by something outside of the person's control, such as a brain injury, a developmental disorder, or features of their childhood. The existence of these constraints on the mental state also reduced subjects' judgments that the holder of the mental state was blameworthy for holding the attitude, or that they should be confronted for holding it. Follow up mediation analyses indicated that these changes were mediated by perceived control over the mental state.

In one sense, these findings are hardly surprising. After all, a great deal of prior work shows that causal explanations for people's illnesses or behavior that cite genetics, biology, or tragic life histories affect perceived control over these phenomena. However, our findings also indicated that these constraints had a lesser, almost negligible effect on non-responsibility moral reactions, such attributions of immoral character and avoidance. Subjects largely thought that the person who felt an immoral attitude, such as liking the idea that his mom would die, had poor moral character and should be avoided, no matter whether that attitude was within or outside his control. And though control predicted character to some degree it did so less than it did blameworthiness. Indeed, it was this differential effect of control on blameworthiness and character that explained why stipulating constraints on people's immoral mental states reduced blameworthiness but not character. Thus, we found additional evidence that control is more important for attributions of moral responsibility than for character, as well as further support for the claim that character is not sufficient or necessary for holding someone morally responsible for an immoral mental state.

General Discussion

Our aim in this chapter was to determine whether people's tendency to hold each other responsible for their mental states, including bad emotions, desires, beliefs, and other attitudes, violates the widely-supported claim that attributions of control are a necessary condition for responsibility. The apparent threat to control stems from two observations. First, many scholars claim that ordinary people judge mental states to be largely involuntary; that is, they are not chosen and there is little people can do to intentionally change them (e.g., D'Andrade, Katz & Postal, 1962; Malle & Knobe, 1997a; Sabini & Silver, 1998). This observation, coupled with the observation that people often seem to hold others responsible for their mental states, appears to counter the idea that control is necessary for moral judgment (Adams, 1984; Smith, 2008). Second, existing research showed that mental states are treated as highly diagnostic of moral character, and therefore affect whom people like or dislike, and with whom they form relationships. Thus, mental states appear to be the clearest case in which an alternative criterion – namely, character diagnosticity – is deemed sufficient for moral responsibility. However, little work directly investigated whether ordinary people attribute control to others over their mental states, or whether control, rather than character, affected their moral evaluations of others' mental states.

Prior to the present inquiry, past research was equivocal regarding whether ordinary people view others' mental states as controllable or uncontrollable. Some evidence had suggested that people do attribute to others moderate agency over their mental states (Schlesinger, 1992), and this view received recent empirical support with respect to one specific mental state category, namely beliefs (Turri et al., 2017).

However, some empirical research suggested that mental states are seen as largely uncontrollable (Gilovich & Regan, 1986; Malle & Knobe, 1997b). Indeed, some social psychologists had theorized that our mental states are so unintentional and uncontrolled that they "just happen" to us (Gilovich & Regan, 1986). However, all prior research in this domain has been limited by sparse and arguably unrepresentative sampling of mental states.

Additionally, little work investigated the role of control in people's reactions to others mental states. Several studies have found that a person's experiencing deviant emotional reactions to harmful or otherwise unwanted events can result in attributions of poor character to that person and a desire to avoid them (e.g., Ames & Johar, 2009; Gromet, et al., 2016; Szczurek et al., 2012). In some cases, deviant emotional reactions can also inflate a desire for punishment: an individual who acts harmfully and subsequently experiences pleasure or indifference is punished more harshly than someone who merely acts harmfully (Gromet et al., 2016). However, with the exception of Cohen and Rozin (2001), prior work on moral judgments of mental states has overlooked an important dimension, namely the extent to which people can control them. None of these studies measured or manipulated the controllability or intentionality of the mental states in question, nor did these studies measure judgments of blame for those mental states. The exception to this, a series of studies reported in Cohen & Rozin (2001), suggest such a link between these constructs, such that greater control over mental states is associated with heightened blame, but again, this investigation considered only a very limited range of stimuli and was largely correlational.

Across three studies we confirmed two important findings. First, people by and large judged others to have a moderate-to-high degree of control over their mental states. Across several different kinds of mental states (including emotions, desires, beliefs, and evaluations) the typical amount of control our subjects attributed was at or above the midpoint of a seven-point scale – ratings that were more similar to those of prototypical intentional acts than to those of prototypical unintentional behaviors (e.g., sneezing) or accidental behaviors (e.g., dropping something). Even emotions, the lowest rated mental state category, were judged as more controllable and intentional than unintentional behaviors (and sometimes more controllable than accidental behaviors). Thus, while we replicated prior work showing that mental states are judged as less controllable and intentional than typical voluntary behaviors (Gilovich & Regan, 1986; Johnson et al., 2004; Malle & Knobe, 1997a), on the basis of our findings no type of mental state could be said to "just happen." These results call into question what we perceive to be the dominant view amongst scholars; namely, that mental states are judged by ordinary people to be passively experienced and basically uncontrollable.

But even if people normally view mental states as somewhat controllable, it was possible that this control played little role in how they attribute responsibility to others. Yet this was not the case. Across all our studies, we found that control strongly correlated with attributions of responsibility and blame. People who tended to think that emotions, beliefs, desires (and so on) are uncontrollable also tended to attribute very low responsibility and blame to the targets in the vignettes. Furthermore, in Study 1.3, control strongly correlated with holding someone responsible for a mental state, in this case, confronting someone and making them feel bad about their attitude. Thus, we have strong evidence against the claim that people hold each other responsible for their mental states despite believing that they cannot control them. Finally, in Study 1.3 we manipulated control by altering the causal history of the immoral mental state. When the immoral attitude was caused by features of someone's biology or past that were outside of their control, then people judged the attitude itself to be outside of the target's control. And, in these conditions, people judged the target as less blameworthy, and were less inclined to confront the target, over the immoral attitude. These findings are the first to our knowledge to establish causal evidence for the role of control in mental state responsibility.

We also observed a positive relationship between perceived control and character judgments. Taken at face value, this positive relationship makes sense – just as we might regard more controlled behaviors as more clearly reflecting a person's character than less controlled behaviors, the more control a person has over an aspect of their mental functioning, the more that aspect of the mind would seem to reflect a deep part of their character. However, some other results suggest otherwise. For instance, a meta-analysis of lay beliefs showed that endorsement of biogenetic explanations for mental illnesses is positively correlated with judgments that those illnesses constitute an essential component of those individuals, and with the desire for social distance from them (Kvaale, Gottdiener, & Haslam, 2013). These responses imply the attribution of bad character, and yet as the authors point out, "In the framework of attribution theory, biogenetic explanations reduce perceptions of the controllability of behavior" (p. 99). Along similar lines, Suhay, Brandt, and Proulx (2017) found that people who believe that political views are biologically caused are more intolerant of, and more avoidant of, those who hold opposing political views. This finding again seems to show that deep character inferences flow from biogenetic explanations, and yet those explanations also imply a lack of control. Notwithstanding these findings, our studies on mental state control yielded a contrasting pattern, such that control over mental states was positively associated with character diagnosticity.

One possible explanation of our findings is that character in fact is the basis for holding others' responsible, and that other judgments of responsibility and control are downstream consequences of this. For instance, perhaps people are motivated to view character-diagnostic attitudes as blameworthy, and further motivated to judge blameworthy conduct as controllable (Alicke, 2000; Nadler & McDonnell, 2012; Pizarro et al., 2013). Or perhaps people judge character-diagnostic attitudes as more intentional or controllable than non-diagnostic attitudes. While our data cannot rule out that character evaluations ever affect attributions of control or moral responsibility, we can rule this explanation out with respect to the findings presented above. First, a motivated account of control is plausible for highly immoral mental states, such as those used in Studies 1.2 and 1.3, but it is not plausible for neutral mental states used in Study 1.1. And yet, in Study 1.1 we observed comparable levels of control attributed to the mental state holders, and control, responsibility, and character were still correlated with one another.

Second, and more importantly, we repeatedly observed a dissociation between character and responsibility such that responsibility appears to require control and character does not. In Studies l.1 and 1.2, this dissociation was evident through the observation that control was much more strongly associated with responsibility (and blameworthiness) than it was with character. This disassociation revealed a pattern of results which strongly hint that control, not character, underlies attributions of mental state responsibility. For instance, in Study 1.1, subjects who on average attributed very low control to others over their mental states tended to also attribute very low responsibility to those individuals over the mental state as well. And yet, these same subjects still associated some degree of character-relevance to the mental state (see Figure 1.3). This suggests not only that control may be a necessary precondition for responsibility, but also that character is not a sufficient precondition. This claim gained further support in Study 1.3. In Study 1.3 we showed that it was possible to causally intervene on perceived mental state control, and that doing so could change attributions of moral responsibility more strongly than attributions of poor moral character. Thus, reducing control by stipulating that someone's mental state was, for instance, caused by a developmental disorder, reduced perceived blame for the mental state but not judgments that the target was a bad person for having that mental state. This again suggests that control is necessary for responsibility – as reducing perceived control reduces perceived responsibility – and that character inference is not sufficient for moral responsibility – as even though people still inferred poor character, responsibility attributions reduced. In total, these results rule out an alternative proposal that character evaluation is the principle determinant of moral responsibility for mental states.

There are two notable limitations to the investigation reported here. First, though our investigation was motivated as a way to test whether control was necessary for moral responsibility when evaluating mental states, our conclusions are limited by the fact that people by-and-large attributed control to others over their mental states. This leaves open the possibility that there is still some highly blameworthy-but-uncontrollable mental state that we failed to investigate, and therefore still a devastating counter-example to the claim that responsibility requires control. Though this is possible, we are pessimistic that such cases naturally occur or could be believably portrayed. Our studies did measure people's reactions to many of the putative examples of involuntary, but sinful, mental states such as "a lack of concern for others," contempt," "hatred," and "anger" (c.f. Adams, 1985) Yet, these were all judged to be at least somewhat controllable, and in some cases highly controllable. For this reason, it is not clear what a good candidate for an 'uncontrollable but highly blameworthy' mental state would be. Moreover, even when it was stipulated that someone had an immoral mental state due to some developmental disorder, people still attributed some control to him over it. Thus, we are pessimistic that, in realistic contexts, people will ever view an immoral mental state as unambiguously outside of someone's control.

Our second limitation stems from the observation that all of our studies were all surveys where, among other things, we provided subjects perfect knowledge of someone's attitude, explicitly prompted subjects to answer questions about control, and provided subjects with little other information about the target and his/her character. Furthermore, people made judgments about sparsely realized characters with whom they have no relationship. To compare: in real life contexts, there is often a great deal of uncertainty about someone's attitude, people are rarely explicitly prompted to reason about the controllability or causal history of someone's mental state, and people are embedded in a social context in which the character of the individual they are making judgments about is much more important. Given all of these differences, it is plausible that, in real life settings, people base responsibility judgments on more salient or personally important information, including information about the target’s character. For these reasons, we believe that an important direction for future work is to collect data regarding how people judge close others’ immoral, objectionable, or hurtful mental states, and test whether even in these highly personal, rich contexts, people still hinge their reactions on judgments that the close other has, or had, voluntary control over the mental state. For these reasons, our findings should not be considered definitive, and await further investigation.

Conclusion

Mental state blame has historically been the context that represents the best case against the putative necessary role of control in moral judgment. This “best case” status stems from three observations. First, mental states are *prima facie* highly related to someone’s character and identity, and therefore may be revealing about someone for reasons other than on the basis of personal control. Second, people appear to blame others for poor and objectionable mental states. And third, mental states were widely considered to be too far outside people’s control to justify moral responsibility. However, despite the apparent usefulness of mental state blame as a way to adjudicate debates about the role of control in blame, no work directly empirically investigated it until now. We presented three studies that show that this challenge lies on a false premise. Contrary to popular scholarly belief, ordinary people largely believe that people can control what they feel, think, and want. And these attributions of control predict when they judge someone responsible for mental states, and when they feel justified to hold someone responsible for their mental state.

Regulating Emotion Regulation

When people feel upset, stressed, or anxious, they reach out to others to help regulate those negative feelings away. In general, the people they reach out to (hereafter, *observers*) feel sympathy for suffering others (hereafter, *sufferers*), are motivated to help them (Batson, 1991) and often succeed in doing so (Williams, Ong, & Zaki, 2018). And yet, in many situations, observers choose not to help, leaving the sufferer to cope with their sadness, anxiety, or distress on their own (Dunkel-Schetter & Skokan, 1990). Moreover, observers also sometimes react toward the sufferer in intentionally unsupportive ways. For instance, they might yell at someone to "get over it," impugn their character, or express irritation at the sufferer for feeling upset. Why might observers sometimes treat others this way? More generally, what predicts whether observers will react in a supportive or unsupportive manner toward others who are suffering from a negative emotion?

One reason why observers may withhold sympathy is because, in that moment, they consider it too costly to take on the burden of making the sufferer happy (Cameron et al., 2019). Helping someone can be costly because observers find it stressful to be around others who feel upset (Cialdini et al., 1997; Coyne et al., 1987). Observers also feel averse to feeling responsible for someone's emotional well-being, which can happen when they agree to help (Coates, Whortman, & Abby, 1979; Wortman & Lehman, 1985). However, an aversion to helping others does not readily explain why observers might act in purposefully unsupportive ways, too. In fact, if people are primarily motivated by an aversion to others' distress, or an aversion to feeling responsible for others' suffering, then they should be highly motivated to avoid making people feel even worse. Therefore, mere considerations of the personal cost of helping seem unlikely to explain the full range of reactions observers may have to others' suffering.

People also withhold help when they believe that the sufferer deserves his or her plight. Like many behaviors, emotional reactions can reflect morally objectionable goals, attitudes, or values, and therefore generate inferences of poor moral character (Ames & Johar, 2009; Gromet et al., 2014; Sabini & Silver, 1999; Szczurek, Monin, & Gross, 2012; Uhlmann et al, 2013). People tend to feel less concerned about, and feel less sympathy for, others with poor character (e.g., Brambilla, Hewstone, & Colucci, 2013; Gromet et al., 2016; Monroe & Platt, 2019). To illustrate this point, observers would not feel sympathy for someone who feels upset about a failed murder but in fact might relish that person's suffering. Observers also judge sufferers as undeserving of sympathy when they perceive that the sufferer is at fault for their bad situation in the first place (Weiner, 1995; Weiner et al., 1988). Thus, if someone is upset about something that is ultimately his own fault – for instance, someone is sad because he doesn't have a job even though it was his decision to quit – we should expect observers to react toward him in an unsupportive manner.

But a deservingness account of emotional support cannot be the whole story, either. Consider the following mundane examples of emotional distress:

- Someone feels nervous before a presentation even though she has diligently practiced her talk and has a history of successful public speaking,

- Someone feels upset about a personal trauma they experienced even though it happened several months ago,
- Someone feels embarrassed about committing a *faux pas* even though no one around seems to think that they have behaved in an untoward way.

In these situations, friends and close others may display irritation towards the sufferer or otherwise be poorly motivated to provide emotional support. And yet, in these situations, the sufferer's emotion does not diagnose them as a bad person nor is due to some reckless past behavior. It seems that people must be basing their decision to be supportive or unsupportive on some other criterion.

The "Regulating Emotion Regulation" Hypothesis

Attending to others' capacity to fix their own problems provides observers a method for balancing competing goals in the context of close relationships. As noted above, people both desire for close others to be happy (and free from suffering), but also generally wish to avoid taking on costly burden or taking on burdens that are unnecessary or not shared with others. If a close other does not need help to improve, then people should be less inclined to offer it because they can still realize one goal (the sufferer feeling better) while simultaneously achieving another (not taking on emotion labor; Batson, 1991, Cialdini, 1987). Past empirical work suggests people react to others in a way consistent with this reasoning. People are much more likely to help out others who have already attempted to help themselves and failed (Karasawa, 1991; Meyer & Mulherin, 1980; see also Dunkel-Schetter & Skokan, 1990, for a review in the context of coping with trauma). We hypothesize that this reasoning about sufferers' capacity to fix their own situation (sometimes referred to "offset-control," Brickman, 1990; Karasawa, 1991) extends to reasoning about others' capacity for emotion regulation. When observers judge that others are suffering from emotions they can regulate away, they will judge those individuals as not in need of help and will offer little sympathy or accommodation.

Attending to whether someone else can regulate their emotions themselves also explains why observers sometimes intentionally act in unsupportive ways. Criticizing someone or expressing frustration toward that person is an effective means to motivating them to change their behavior (see, e.g., Malle, Guglielmo, & Monroe, 2014, for discussion of the social function of blame). For this reason, people may believe they can minimize how costly it is to themselves to see a sufferer's emotion reduced by criticizing or reacting unsupportively towards a sufferer. Because unsupportive behavior is a means to an end to achieving less suffering *on net,* it explains why people may intentionally behave unsupportively despite a general aversion to causing others' distress. But, if this logic guides people's behavior, then observers should refrain from unsupportive behavior when someone lacks the capacity to regulate away their suffering on their own. This is because, in these situations, negative behavior causes harm but without bringing about any downstream good effects, and therefore fails to achieve the goals of the observer (Heider, 1958; Brickman et al., 1982; Karasawa, 1991).

This line of researching predicts that observers' supportive and unsupportive behavior should be strongly determined by their prior attributions of emotion control (Figure 2.1). When observers attribute high emotion control to sufferers, they will withhold supportive behaviors and react in purposefully unsupportive ways. But, when observers judge that sufferers have little control over their emotion, they will offer sympathy and accommodation. We refer to proposal as the “Regulating Emotion Regulation” (RER) hypothesis. Below we motivate the central pillar our proposal – that people regularly attribute emotion control to others – and then outline a strategy for predicting how attributions of control, and therefore supportive and unsupportive behaviors, vary across people and situations.

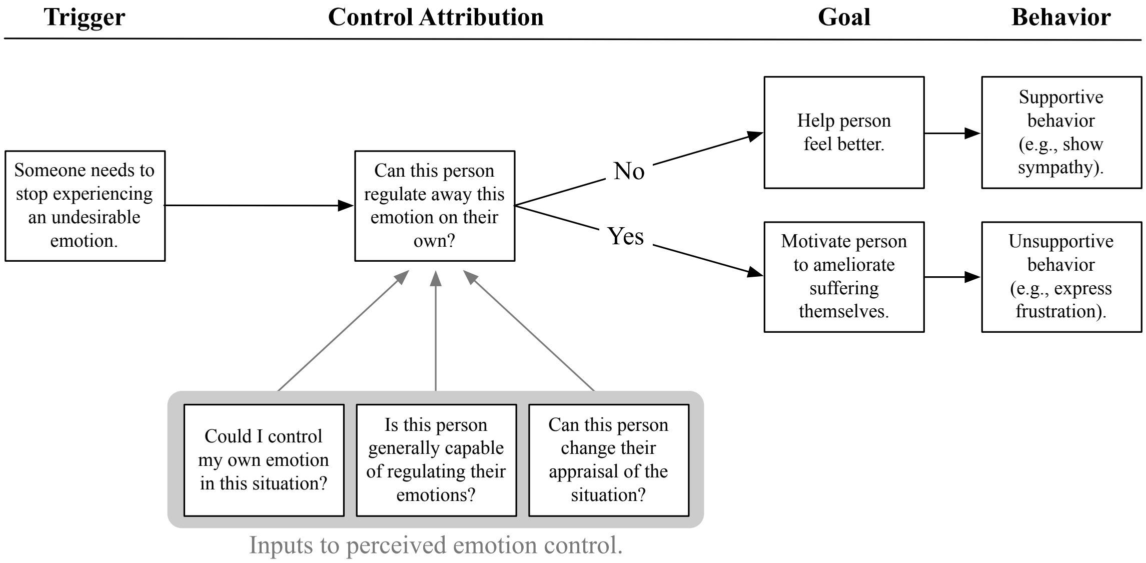
Trigger
Control Attribution
Goal
Behavior
Someone needs to stop experiencing an undesirable emotion.
Can this person regulate away this emotion on their own?
No
Yes
Help person feel better.
Motivate person to ameliorate suffering themselves.
Supportive behavior (e.g., show sympathy).
Unsupportive behavior (e.g., express frustration).
Could I control my own emotion in this situation?
Is this person generally capable of regulating their emotions?
Can this person change their appraisal of the situation?
Inputs to perceived emotion control.

Figure 2.1. The Regulating Emotion Regulation (RER) hypothesis.

Reasoning about others’ emotion regulation control

People play an active role in determining what emotions they experience and how strong they experience those emotions. When individuals feel sad, frightened, angry, or distressed, they have a suite of cognitive and behavioral tools at their disposal to reduce those undesirable feelings in favor of desirable ones (Gross, 1998; Koole, 2009; Parrott, 1993; Tamir, 2009). Some techniques rely on other people for help (i.e., *inter*personal emotion regulation, Williams, Ong, & Zaki, 2018), but many do not. *Intra*personal strategies for regulating emotions include distracting oneself from the source, suppressing negative thoughts, thinking pleasurable or relaxing thoughts, and reappraising the situation, among others (Derakshan et al., 2007, Fernandez, 1986; Ochsner & Gross, 2008; Van Dillen & Koole, 2007; Wenzlaff & Wegner, 2000). These intrapersonal mechanisms are effective: people who use them more (e.g., because they are more motivated to) tend to recover from trauma more quickly (e.g., Tamir et al., 2007; Ford et al., 2019). If people recognize that others are capable of executing these intrapersonal strategies of emotion regulation, they may judge that others have control over (and are sometimes responsible for) their negative emotions.

Indeed, recent experimental work shows that observers attribute a moderate degree of control to others over what emotions those individuals feel, and these attributions of control influence how observers behave (Chapter 1; Ford & Gross, 2019; Halberstadt et al, 2013; Tullett and Placks, 2016). In Chapter 1, we reported that people judge others to have some intentional control over mundane emotional experiences and that the degree of control people attributed correlated with attributions of responsibility toward that person for that emotion. These attributions of control and responsibility can have important downstream consequences on behavior. For instance, parents who think emotions are controllable report being less supportive of their child's negative emotions and more likely to express irritation or behave in punitive ways (e.g., sending a child to his or her room; Halberstadt et al, 2013).

However, prior work on perceived emotion control is limited in several ways. Most notably, no work to date has established causal evidence for the impact of perceived emotion control on supportive and unsupportive reactions. Prior work which showed an association between control and responsibility, or control and sympathy and helping behavior, relied exclusively on correlational evidence. In fact, to date the evidence suggests that there is no causal link between perceived emotion control and people's reactions to others' emotions. For instance, Tullett and Placks (2016) report that across four studies they were unable to reliably detect a causal association between believing that happiness is controllable and people's sympathy for people who feel unhappy. Thus, one major contribution of the present work is to provide a method for reliably manipulating perceived emotion control, and in a way that establishes a causal link between control and downstream supportive reactions.

A second limitation of prior work is that it provides no basis for predicting *situational* variation in perceived emotion control. Past work has relied almost entirely on how perceived emotion control varies across individuals. This leaves no basis for predicting when, or explaining why, an individual might attribute to his friend low control over her emotion in one situation (and offer her sympathy), but attribute to her high control over a similar emotion in another situation (and get angry at her for it). To gain traction on both these problems, we propose a modest starting point: assume that people's attributions of control are at least partially sensitive to which kinds of emotional experiences are easier, or more difficult, for people to regulate. In this paper we explore one potential source of variation in emotion control: what degree observers judge that someone's emotion is calibrated to his or her situation.

One way people can sometimes exercise control over their emotions is through cognitive reappraisal. People's emotional reactions are often the result of their appraisal of a situation – that is, their belief that something good or bad is happening and how important that good or bad thing is to them (Lazarus, 1991). For this reason, people can change their reaction by reinterpreting the situation or reprioritize their goals (Beck et al., 2001; Gross, 1998; Ochsner, Bunge, Gross, & Gabrieli, 2002; Simon, Greenberg, & Brehm, 1995; Wilson, Gilbert, Centerbar, 2003). However, appraisals differ in how malleable they are across different situations. One feature that predicts actual appraisal malleability appears to be how calibrated that appraisal is to the situation it is about (Troy, Shallcroft, & Mauss, 2013; Troy, Ford, McRae, Zarollia, & Mauss, 2017; Suri et al., 2018). For instance, if someone feels stressed about something that she has no control over, then she can reappraise the stress away by realizing that she has no control over her outcome (and therefore that her stress is not appropriate to the situation). However, if she really is responsible for preventing a terrible thing from happening, then her stress is properly calibrated to her situation and trying to reappraise her stress away will be extremely difficult (e.g., Troy, et al, 2013; Troy et al., 2017).

We hypothesized that lay people will attribute control to others over emotions in part by reasoning about how calibrated a person's emotion is in that situation. Past work lends credence to this hypothesis. People tend to spontaneously attribute beliefs and desires to others when reasoning about their emotions (Ong, Zaki, & Goodman, 2015; Ellsworth & Scherer, 2003; Sabini & Silver, 1998; Saxe & Houlihan, 2017). Additionally, other work suggests that people tend to judge beliefs and desires that they judge as miscalibrated as more malleable and controllable than calibrated ones (e.g., Chapter 1, Chapter 3; see also, Rogers, Moore, & Norton, 2017). Therefore, if people judge miscalibrated appraisals to be more easily changed than calibrated appraisals, then they should judge the emotions that flow downstream from those appraisals as more controllable, too. Then, according to the RER hypothesis, people will be less sympathetic toward others, and more likely to criticize that person for having the negative emotion. In the studies below, we manipulate emotion calibration in order to modify attributions of emotion control and these changes in perceived control should then influence people's subsequent supportive and unsupportive behavior.

The current studies

Overview of hypotheses.

To summarize, we hypothesize that people sometimes opt to withhold sympathy in favor of regulating someone else's intrapersonal emotion regulation when they judge that those individuals are capable of regulating their negative emotions away on their own (Figure 2.1). Across six studies we test five specific predictions that fall out of this theory. Two hypotheses make up the central tenants of the regulating emotion regulation hypothesis:

Hypothesis 1: When a sufferer is seen as having the ability to make themselves stop feeling a bad emotion, observers will feel and offer less sympathy to that person.

Hypothesis 2: When a sufferer is judged as having the ability to make themselves stop feeling a bad emotion, observers will report an intention to treat that person in an unsupportive manner for feeling that negative emotion.

Hypotheses 3 and 4 reflect possible cues that people attend to when attributing the capacity to regulate a negative emotion away.

Hypothesis 3: Observers who believe they are capable of controlling their own emotions should judge that sufferers are more capable of control their own emotions as well. As a result, observers should show less sympathy (and more hostility) towards sufferers.

Hypothesis 4: Emotions caused by miscalibrated appraisals will be judged as more controllable than emotions which stem from calibrated appraisals. This is turn results in miscalibrated emotions receiving less supportive, and more unsupportive reactions from others.

And finally, hypothesis 5 states that miscalibration is only an effective cue of control in sufferers who are rational and clear-headed. Due to this, miscalibration does not directly affect sympathy and unsupportive behavior, it only does so indirectly through perceived control.

Hypothesis 5: Perceived emotion miscalibration leads to higher attributions of emotion control, and subsequent unsupportive behavior, only sufferers who are judged to be have normal functioning emotion regulation capacity.

Overview of studies.

In Study 2.1, we investigate whether individual differences in perceived emotion control predicts supportive and unsupportive reactions to other's negative emotions. In Studies 2.2 – 2.3, we experimentally vary whether a target has a calibrated emotion or a miscalibrated emotion, and test whether calibration predicts perceived control, supportive behavior, and unsupportive behavior. In Study 2.4, we address a worry that subjects' reactions are directly in response the miscalibration of the emotion, rather than the control that miscalibration cues. In Studies 2.5 and 2.6 we show that the RER hypothesis can be applied to predict and explain people's behavior in two important contexts. In Study 2.5, we report an autobiographical recall study in which subjects recalled and wrote about a recent time someone close to them experienced a negative emotion. And finally, in Study 2.6, we apply the RER hypothesis outside the domain of close interpersonal relationships and show that attitudes about the controllability of emotions predict support for anti-microaggression University policies (e.g., safe spaces, bans on offensive behavior). Thus, across all studies, we provide support for our theory that people's decision to feel sympathy and help, or to feel irritated and criticize, in response to an emotion is driven by their assessment that the person in question can regulate away that negative emotion him or herself.

Study 2.1

Subjects in Study 2.1 read four vignettes about a close friend feeling upset, embarrassed, stressed, and distressed and reported their reactions. Our primary interest was whether individual differences in perceived emotion control across these vignettes correlated with supportive and unsupportive reactions (H1-H2). As a secondary goal, we hypothesized that people who found it easier to regulate their emotions would judge others as also more able to regulate their own, and that this would influence how they react to others (H3).

Methods

Participants. 190 subjects (100 Female, mean Age = 32.3) were recruited from a University-administered data collection panel. This sample size reflects one recruitment session, from which we expected to recruit 150-200 subjects. We preregistered that we would stop data collection after one recruitment session. No data were discarded. Our final sample size had a 99% chance to detect a small effect (r = .3) or larger.

Experiment context. Subjects filled out two surveys that ostensibly had nothing to do with one another except that both were included in a set of studies as part of a University-administered data collection panel. In one survey, subjects reported their reactions to four vignettes in which a close friend has a negative emotional reaction. In the second survey, subjects believed they were responding to a survey about the relationship between attitudes and life satisfaction and, as a part of this survey, reported how much control they believe they have over their emotions. These two surveys were separated by another experiment completely unrelated to emotion regulation, moral judgment, or sympathy. See Appendix E for scales used in this survey. We obtained age and sex information from a separate demographics survey collected in the same experiment session.

Vignette Task. In survey one, subjects read and reacted to four vignettes that described someone having a negative emotional experience. For instance, one vignette described someone hiking, barely hurting their leg, and then having a severe reaction of distress in response to the injury. The other three vignettes described someone feeling sad about receiving an A- on an exam, feeling embarrassed about a very minor flaw in a cake, and feeling stressed about giving a short presentation in class. Though the type of emotional reaction varied across vignettes, each vignette shared certain features. First, the emotion in each vignette was described as an intense emotion. Second, the emotional reaction was moderately inconvenient for the subject. Third, each emotional reaction occurred directly following some event (i.e., the target had just fallen and hurt his leg). And finally, the target in each vignette was described as being a close friend to the subject. See Appendix F for full text of each vignette.

Subjects responded to four questions for each vignette. These included (1) how much control the target has over the emotion (e.g., "If he wanted to, Arthur could choose to stop feeling distressed."), (2) how much sympathy the subject feels (e.g., "I would feel a great deal of sympathy for Arthur."), and (3) whether they would make the person feel bad (e.g., "I would make Arthur feel bad for feeling this distressed."). This latter item was our measure of unsupportive reactions. Subjects responded to these questions using a 7-point agreement scale (1: strongly disagree, 7: strongly agree). A fourth question measured the perceived strength of the target's emotion (e.g., "How much distress is Arthur experiencing?" 1: none at all, 7: a great deal of distress). The name of the target and emotion label within each question varied to match the vignette.

Emotion theory task. In survey two, subjects filled out a short survey about life satisfaction. In the first section of this survey, subjects responded to a Life Satisfaction Scale (Diener, Emmons, Larsen, & Griffin, 1985), followed by an Implicit Beliefs about Intelligence Scale (Dweck, 1999). Subjects then filled out the Implicit Beliefs about Emotion Scale (DeCastella et al., 2013), indicating their agreement (1: strongly disagree, 5: strongly agree) with four statements written in the first person about emotion control: "If I want to, I can change the emotions that I have," "I can learn to control my emotions," "The truth is, I have very little control over my emotions," (R) and "No matter how hard I try, I can't really change the emotions that I have" (R).

Results

As planned, we averaged each subjects' control (α = .73), sympathy (α = .68), unsupportive reactions (α = .70), and emotion strength (α = .66) judgments across the four vignettes. As predicted, subjects' judgments of emotion control negatively correlated with the degree of sympathy they felt toward the target, *r*(188) = -0.44, *p* < 0.001, 95% CI [-0.55, -0.32], and positively correlated with whether they would make the target feel bad for their emotional reaction, *r*(188) = 0.45, *p* < 0.001, 95% CI [0.33, 0.55] (Table 2.1; Figure 2.1A-B). We also observed that people who perceived the emotions as being stronger showed more sympathy, *r*(188) = 0.28, *p* < 0.001, 95% CI: [0.14, 0.40], and were less likely to report making the person feel bad, *r*(188) = -0.18, *p* = 0.01, 95% CI: [-0.32, -0.04]. However, there was no significant association between emotion strength and control, *r(*188) = -0.11, *p* = 0.13, 95% CI [-0.25, 0.03].

Table 2.1
Means (and standard deviations) and correlations between

		Correlation Coefficient			
Variable	**Mean (SD)**	**2**	**3**	**4**	**5**
1. Perceived Emotion Control	4.12 (1.30)	0.45**	-0.44**	-0.11	0.29**
2. Unsupportive reaction	2.58 (1.21)		-0.41**	-0.18*	0.06
3. Sympathy	3.92 (1.23)			0.28**	-0.17*
4. Emotion Strength	5.55 (0.87)				.02
5. 1st Person Emotion Control	3.55 (0.82)				

Note. Ratings made on a 1-7 scale for questions 1-4, and a 1-5 scale for item 5.
* *p* < 0.05, ** *p* < 0.001

We next analyzed the relationship between subjects' judgments about how much control they have over their own emotions and their reaction to the targets in the vignette. We averaged subjects' responses to the Implicit Beliefs about Emotion Scale (*alpha* =

0.76), and then measured its association with subjects' average control, sympathy, unsupportive reactions, and strength judgments. As expected, subjects who thought they were more capable of controlling their own emotions tended to attribute more control to others, *r*(188) = 0.29, *p* < 0.001, 95% CI [0.15, 0.41] (Figure 2.2C). Notably, they did this even though they did not perceive the intensity of others' emotions any differently than those who judge their own emotions as highly uncontrollable, *r*(188) = 0.02, *p* = 0.83, 95% CI [-0.13, 0.16].

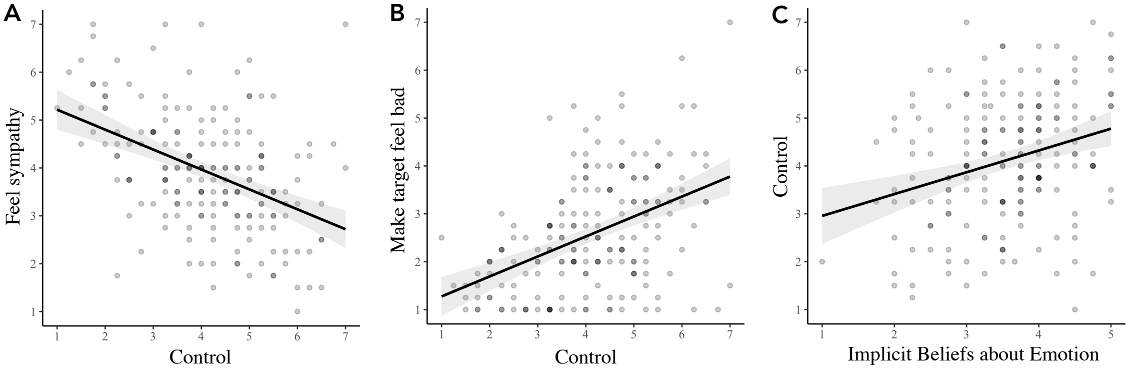

Figure 2.2. Key findings from Study 2.1. (A) Subjects' control attributions strongly negatively predicted their average sympathy reactions (values averaged over four scenarios). (B) Subjects' control attributions strongly positively predicted their punishment reactions (values averaged over four scenarios). (C) Individual differences in average control attributions to the vignettes was positively correlated with individual differences in judgments that others have control over their emotions.

Subjects implicit beliefs about their own emotion control affected how they reacted to others. Subjects who reported higher emotion self-control tended to feel less sympathy for the targets in the vignettes, *r*(188) = -0.17, *p* = 0.02, 95% CI [-0.30, -0.03].

As expected, differences in control attributions towards the targets in the vignettes mediated the relationship between self-attributed control and sympathy (*a* = 0.46, *p* < .001, *b* = -0.40, *p* < .001, *ab* = -0.18, 95% CI [-0.30, -0.09])[2]. As planned, we repeated this analysis on unsupportive reactions. Unlike sympathy, implicit theories of emotion did not correlate with unsupportive reactions, *r*(188) = 0.06, *p* = 0.38, 95% CI [-0.08, 0.20]. However, there was still an indirect effect of control mediating subjects' implicit theories of belief and their unsupportive reactions (*a* = 0.46, *p* < .001, *b* = 0.44, *p* < .001, *ab* = 0.20, 95% CI [0.10, 0.32]). Thus, self-directed control beliefs are associated less sympathetic reactions to others' suffering, because people who tend to think their own emotions are controllable also tend to judge others' emotions as controllable, too.

Discussion

Consistent with the RER hypothesis, we observed that individual variation in perceived emotion control predicted the degree of sympathy that those individuals feel for the sufferer, as well as the likelihood that the person will choose to add to that person's suffering. The more control that subjects attributed to the target, the less likely they were to feel bad for the target, and the more likely they were to make the target feel bad for feeling bad. We also observed that one source of individual variation in perceived

[2] We follow recommendations from Yzerbyt, Muller, Batailler, & Judd (2019) for conducting and reporting mediation analyses. When reporting mediation analyses, we first results from the joint significance tests of the a-component (a) and b-component (b) of the mediation model and conclude that there is mediation when both a and b are significant. We then report the boot-strapped estimated size of the indirect effect (*ab*) and its 95% confidence interval. With the exception of our multiple mediation models reported in the Appendices, all mediation and moderation analyses were carried out using the 'JSMediation' package provided by Yzerbyt et al (2019).

control stems from subjects' sense that they can control their emotions themselves. Though it is possible that this association is partially caused through demand or a desire for consistency, we took multiple steps to reduce these effects. First, we separated our trait measure from our vignettes across surveys. Additionally, although the lay theories of emotion scale asked about "control" explicitly, our measure of control in the vignette study was indirect, asking instead whether the target could change her emotion if she wanted. Thus, it appears one source of support towards others is one's own sense of emotion control.

Although results from Study 2.1 are consistent with the RER hypothesis, they are only correlational and take place in the limited context of a hypothetical vignette (with a hypothetical close friend). We address these limitations in the following studies. In Studies 2.2 – 2.4, we manipulate perceived emotion control by manipulating how calibrated or miscalibrated the emotion is. In Study 2.5, we measure people's real-life supportive and unsupportive reactions in an autobiographical recall study. And in Study 2.6, we show that attributions of control predict people's attitudes in a policy context – namely, their attitudes about what Universities ought to do to protect minorities from microaggressions.

Study 2.2

It makes sense for someone to be extremely upset about their car being stolen, as losing a car is significant personal loss. But it would seem irrational for someone to feel equally upset over a stranger's car being stolen. The latter would suggest that someone's subjective assessment of the event's seriousness is wrong, irrational, or otherwise misguided. As we argued in the Introduction, people's attitudes are generally more malleable when they are miscalibrated because they can be changed by exposure to new information or clearer thinking. We hypothesized that people would be sensitive to this feature of emotion regulation, and so would judge miscalibrated emotions as more controllable than calibrated emotions. To test this, we wrote a series of vignettes which varied the intensity or personal relevance of an emotion-eliciting situation but kept constant the perceived intensity of the target's emotional reaction to this situation. We hypothesized that people would judge strong emotional reactions to severe, personal events to be calibrated but equally strong emotional reactions to less severe, less personal events to be miscalibrated. We further predicted that calibration would be associated with perceived emotion control: people would judge that others have more control over miscalibrated emotions compared to calibrated emotions.

Methods

Participants. We recruited 120 people from Amazon's Mechanical Turk (47 Female, mean Age = 37.1). This resulted in 90% power to detect an effect size of d = .3 or greater.

Materials and Procedure. We constructed six scenarios. In one scenario, a hypothetical friend Jamie is upset because her mom is in the hospital following a suicide attempt – something highly personally relevant – or is equally upset because the National Institute of Health has released a report saying that the US suicide rate has increased 1.5% in the past year – something not personally relevant. The other five vignettes described someone feeling upset in relation to (2) brother forgetting birthday vs half-birthday, (3) rain destroying expensive electronic equipment vs sudoku books, (4) significant other getting dinner with an ex vs a work colleague, (5) receiving a C- on an exam vs an A-, and (6) spilling a beer all over oneself vs spilling a beer on the counter. See Appendix G for full text of each scenario. We constructed two lists which each contained three vignettes with the miscalibrated emotional condition, and three vignettes with the calibrated emotional reactions. Subjects were randomly assigned to one of the two lists at the beginning of the study.

Subjects made three judgments in response to each vignette: perceived calibration, perceived emotion control, and perceived emotion strength. To measure calibration and control, subjects rated their agreement with the statements “X’s emotion appropriately matches the situation” and “If X wanted to, he could choose to stop feeling upset,” respectively (where “X” was replaced with the target’s name in the vignette and pronouns matched). To measure perceived emotion strength, subjects responded to the question “How strong is X’s emotion?” on a 7-point rating scale (1: not at all strong, 7: extremely strong). These questions were presented in a random order across the vignettes, and the vignettes were shown in a random order.

Results and Discussion.

We computed subjects’ average ratings for each of our measures within each of the two conditions. All analyses reported below are based on subject averages. As expected, subjects judged the target’s strong emotions in the weakly stimulating condition as less calibrated ($M = 3.34$, $SD = 1.56$) compared to the strongly stimulating condition ($M = 5.01$, $SD = 0.94$), $t(119) = 10.57$, $p < 0.001$, 95% CI [1.31, 1.91], $d =$

1.24. Subjects also judged the target as having more emotion control in the weakly stimulating condition (*M* = 5.10, *SD* = 1.46) relative to the strongly stimulating condition (*M* = 4.20, *SD* = 1.38), *t*(119) = -6.42, *p* < 0.001, 95% CI: [-1.18, -0.62], *d* = -0.63. However, we did not observe a significant difference in perceived emotion strength between conditions: ratings of emotion strength were nearly identical in the weak (*M* = 5.44, *SD* = 0.87) and strong (*M* = 5.60, *SD* = 0.81) conditions, *t*(119) = 1.92, *p* = 0.057, 95% CI [0.00, 0.31], *d* = 0.18. Calibration was negatively correlated with control such that, as subjects judged the emotional reaction to more appropriately match the situation, they attributed to the target less control over feeling upset, *r*(238) = -0.31, *p* < 0.001, 95% CI [-0.42, -0.19]. A follow-up within-subjects mediation analysis showed that the effect of stimulus strength on control was fully mediated by perceived calibration (*a* = -1.61, *p* < 0.001; b = -0.62, *p* < 0.001; c = 0.90, *p* < 0.001; c' = -0.09; *p* = 0.527, *ab* = 0.99, 95 CI [0.73, 1.28]).

Study 2.2 supported our prediction (H3) that people would judge miscalibrated emotions as more controllable than calibrated emotions. Consistent with our proposal in the Introduction, it appears that people judge that others are more capable of changing their mind about something when they perceive that person to have an irrational or wrong attitude. This is consistent with work showing that people believe others' attitudes will converge with their own over time (Rogers et al., 2017) and that people judge irrational or non-normative attitudes are more controllable than normative ones (Chapter 3). However, to our knowledge it is the first direct experimental evidence showing that perceive irrationality affects perceived attitude control. In Studies 2.3-2.4 we tested whether these differences in control predict downstream sympathetic and unsympathetic behaviors.

Study 2.3

Our primary goal in Study 2.3 was to test whether changes in perceived control predicted sympathetic and unsympathetic behavior. Recall that the RER hypothesis predicts that people react unsympathetically toward others in order to motivate them to regulate their emotion themselves. We obtained some evidence for this in Study 2.1, however, the measure we used (agreement with the statement "I would make [this person] feel bad for feeling this [emotion].") is an imperfect gauge of the goal to motivate the sufferer. After all, people may make others feel bad for a wide variety of reasons, including because they enjoy it, or plausibly in many cases, as an act of retribution for some offensive or improper conduct. In order to best measure behavior regulation motives, we require an anti-social, punish-like behavior that preferentially reveals a desire to modify someone's conduct or cognition.

To this end, we recruited 60 people on Amazon's Mechanical Turk for a short task on moral language. We provided them with ten anti-social behaviors and asked them to indicate what typically motivates them to engage in those behaviors. The ten behaviors included *attack*, *yell at*, *criticize*, *express frustration, blame, punish*, *insult*, *ignore*, *avoid*, and *make someone feel bad*. For each behavior, participants selected up to ten reasons they would engage in that behavior. Two reasons described motivations to modify someone's behavior ("get them to change their behavior" and "change how they think about something"), two described retributive motives ("get back at them for something" and "fix an injustice/right a wrong"), two describe motives to increase social distance from someone ("make it clear to them that I do not like them right now" and "get them to stay away from me "), and lastly, two reasons described selfish motives ("make myself feel better" and "manipulate them into doing something for me").

This study revealed that two unsupportive behaviors, *criticizing*, and *expressing frustration*, were especially good candidates for measuring motives to change behavior. For "criticize," 83% of subjects selected at least one *modify behavior* goal, whereas the next highest goal, *retribution*, was selected by only 47% of subjects (note that subjects could select more than one goal). Similarly, for "expressing frustration" a majority of subjects (67%) selected a behavior modification goal, while only 48% selected the next most frequent goal (retribution). None of the other unsupportive behaviors clearly favored behavior modification as a principal motivation. This included "making someone feel bad," which people equally associated with behavior modification and retributive goals, and which we used in Study 2.1. For our remaining studies, starting with Study 2.3, we measured unsupportive behaviors using *criticize* and *express frustration*. Having identified a useful measure of unsupportive reactions, we now turn to testing whether tendency to act in an unsupportive manner can be manipulated by changing how calibrated, and therefore controllable, the emotion is.

Methods

Participants. We recruited 210 people (109 Female, mean Age = 35.1) from Amazon's Mechanical Turk to participate in the main task for Study 2.3. This sample size yielded 90% to detect an effect size of d = .4 or greater.

Procedure. Subjects read a scenario in which a hypothetical friend feels extremely upset either because his/her mom is in the hospital (calibrated condition), or because of a negative suicide statistic (miscalibrated condition). At the beginning of the study, subjects were randomly assigned to either the calibrated or miscalibrated condition. As in Study 2.2, if subjects reported being male, Jamie was described as male in the vignette (i.e., used male pronouns), otherwise Jamie was described as female. See Appendix H for full text of vignette.

Subjects reported six judgments in reaction to the vignette. For five items, subjects reported their agreement or disagreement with a statement about their reaction to Jamie on a 7-point scale (1: strongly disagree, 7: strongly agree). One item measured perceived emotion control ("If Jamie wanted to, she could choose to stop feeling upset."). Two items measured unsupportive reactions ("I would criticize Jamie for feeling this upset" and "I would express frustration toward Jamie for feeling this upset"), which were based on our pretest. And two more items measured supportive reactions ("I would feel a great deal of sympathy for Jamie" and "I would do everything I could to accommodate Jamie"). Finally, as in Study 2.2, we measured perceived emotion strength by asking subjects "How strong is Jamie's emotion?" and providing a 7-point scale anchored at 1 ("not at all strong") and 7 ("extremely strong"). The order of each question was randomly determined for each subject.

Results

We replicated results from Study 2.2 showing that emotion calibration strongly affected perceived emotion control. Subjects attributed more emotion control in the miscalibrated condition (*M* = 4.44, *SD* = 1.82) compared to the calibrated condition (*M* = 2.96, *SD* = 1.64), *t*(203.81) = -6.18, 95% CI [-1.95, -1.01], *p* < 0.001, *d* = 0.86. We also observed a smaller, but still significant difference in perceived emotion strength: Jamie's emotion was seen as slightly weaker in the miscalibrated condition (*M* = 5.78, *SD* = 1.21) compared to the calibrated condition (*M* = 6.19, *SD* = 0.96), *t*(195.78) = 2.74, 95% CI [0.11, 0.71], *p* = 0.006 , *d* = -0.30. See Figure 2.3, below.

We next created composite measures of supportive (*r* = .78) and unsupportive (*r* = .75) reactions by averaging together subjects' responses to the two supportive and unsupportive items respectively. As predicted, subjects were less supportive toward Jamie in the miscalibrated condition (*M* = 4.13, *SD* = 1.73) compared to the calibrated condition (*M* = 5.95, *SD* = 1.07), *t*(172.1) = -9.11, 95% CI [1.42, 2.21], *p* < 0.001, *d* = -1.26. Additionally, and as predicted, subjects were more unsupportive toward Jamie in the miscalibrated condition (*M* = 3.25, *SD* = 1.65) compared to reporting practically no unsupportive behavior in the calibrated condition (*M* = 1.74, *SD* = 1.15), *t*(184.13) = -7.69, 95% CI [-1.90, -1.13], *p* < 0.001, *d* = 1.07. Subjects appeared to trade off supportive reactions in favor of unsupportive reactions, which negatively correlated with one another, *r*(206) = -0.46, *p* < 0.001, 95% CI [-0.56, -0.34].

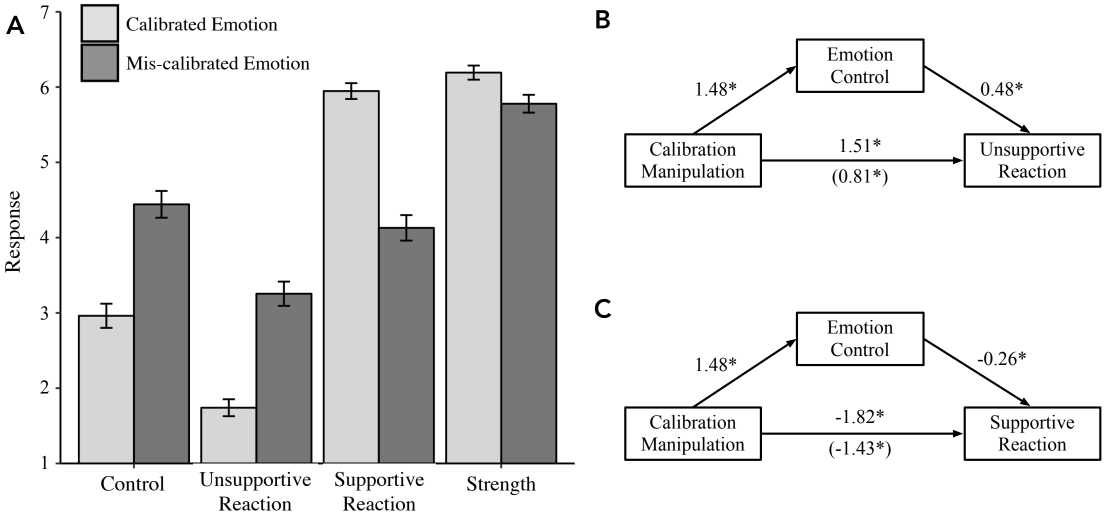

Figure 2.3. Key results from Study 2.3. (A) Means (and standard errors) for each of our measures in Study 2.4 across conditions. (B) Mediation analysis showing significant indirect effect of control on punishment. (C) Mediation analysis showing significant indirect effect of control on sympathy.

One reason that subjects were less supportive and more unsupportive in the miscalibrated condition was because they judged Jamie to have more control over his/her negative emotion. We observed that control strongly correlated with both supportive, *r*(206) = -0.46, *p* < 0.001, 95% CI [-0.56, -0.34], and unsupportive reactions, *r*(206) = 0.66, *p* < 0.001, 95% CI [0.57, 0.73]. Furthermore, planned mediation analyses showed that perceived emotion control partially mediated the effect of elicitation strength on supportive behavior (*a* = 1.48, *p* < 0.001; *b* = -0.26, *p* < 0.001, *ab* = 0.39, 95 CI [0.21, 0.61]) and unsupportive behavior (*a* = 1.48, *p* < 0.001; *b* = 0.48, *p* < 0.001, *ab* = -0.70, 95 CI [-0.98, -0.47]).

Discussion

Results from Study 2.3 were consistent with predictions derived from the RER hypothesis. In this case, we found that people are less supportive, and more unsupportive, toward someone experiencing negative emotions that they judge to be miscalibrated. While Study 2.2 confirmed that the mismatch between a situation and an emotion predicts increased emotion control through judgments of miscalibration, Study 2.3 showed that these differences in control had downstream consequences on supportive and unsupportive behavior. Therefore, one reason why people may show a lack of support to others in many contexts is because they view the emotion to be an ill-fit, or unjustified by the situation, and therefore something the target can deal with his or herself[3].

[3] There is a worry about our measure of control that we have not yet addressed. We have theorized that people will judge miscalibrated emotions as more controllable because they are more cognitively re-appraisable. However, our measure of control does not specifically measure ability to reappraise. In discussion of this research, several people expressed an interest in whether subjects were trying monitor and regulate other's cognitive (i.e., reappraisal) or behavioral (i.e., suppression) manifestations of emotion. We addressed this in a pre-registered study (n = 397) where we conducted a replication of Study 2.3 while asking additional questions that measured capacity to reappraise ("*Jamie can choose to change the way he thinks about this situation*," "*Jamie can choose to think about this situation in a way that calms him down"*) or suppress ("*Jamie can choose to not express how upset he is," "Jamie can choose to keep his emotions to himself"*) the emotion. Consistent with our expectations, subjects judged miscalibrated emotions to be more reappraise-able than calibrated emotions, $t(392.53) = -3.23$, $p = 0.001$. Additionally, reappraisal capacity more strongly correlated with our measure of control, $r(395) = 0.69$, $p < 0.001$, than did suppression, $r(395) = 0.56$, $p < 0.001$ (diff: $b = -0.21$, $se = 0.07$, $p = 0.003$). This suggests that our measure of emotion control in Studies 2.1-2.6 more closely elicits judgments of cognitive control over emotion as opposed to suppressive control. And finally, reappraisal control correlated with supportive behavior, $r(395) = -0.25$, $p < 0.001$, and unsupportive behavior, $r(395) = 0.23$, $p < 0.001$.

We also found something we did not expect. Judgments of suppress-ability and reappraise-ability were highly correlated, $r(395) = 0.63$, $p < 0.001$, and subjects judged miscalibrated emotions to be more suppressible than calibrated emotions, $t(389.82) = -2.66$, $p = 0.008$. We interpreted this finding as suggesting that one way to successfully hide an emotion is to reappraise it (or otherwise reduce the extent to which you feel it).

However, there is an alternative explanation for our findings. Perhaps subjects' degree of support is not a product of their reasoning about the controllability of an emotion, but instead for some other reason associated with the emotion's miscalibration. For instance, people's reactions to miscalibrated emotions may be similar to how they react to immoral emotions – by judging that the person in question is now simply less sympathetic by virtue of having a non-normative emotional reaction. To adjudicate between this proposal, and the RER hypothesis, we need to dissociate emotion miscalibration and control. According to the RER hypothesis, subjects should refrain from punishing someone who has an irrational emotion if there are extant reasons to think that that person nevertheless does not have control over it. By contrast, a model of that eschews considerations of control predicts that people will be unsupportive of others with inappropriate emotions even if those individuals do not have the ability to regulate those emotions away. We pit these two theories against each other in Study 2.4.

Study 2.4

Miscalibrated emotions engender control because the person in question is capable of rationally reappraising the situation. An essential part of this connection between calibration and control is an assumption that the person in question is capable of reasoning clearly and objectively about the situation – that is, that his or her mind will

Consistent with this, in follow-up studies, we found that people rated behaviors like "stopping crying" or "engaging in conversation with others" to be hard to do without "making yourself feel less upset first". Thus, while reappraisal and suppression are conceptually distinct, especially in theoretical models of emotion regulation and people's dispositional emotion regulation strategies (e.g., Gross & John, 2003), it appears to be that the ability to successfully execute one or the other runs together in observer's lay judgments of others.

change when that person attends to and appreciates the reasons why their original appraisal of the situation was wrong. If that person is incapable of reasoning clearly, or there is some other reason why they cannot cognitively change their emotion, then, for that individual, people should not judge that he or she have more control over a miscalibrated emotion relative to a calibrated one. If people's decision to act in a supportive or unsupportive manner is the product of their attributions of control, then people should not be less supportive of an irrational person for his or her miscalibrated emotion. In statistical terms, RER theory predicts that the capacity to think rationally moderates the link between emotion miscalibration, control, and supportive/unsupportive reactions. We tested this moderated-mediation model in Study 2.4. If confirmed, this finding would also rule out the alternative hypothesis that people are less supportive towards others for some alternative reason related to the emotion's miscalibration. Therefore, an alternative prediction is that people will be less supportive (and more unsupportive) toward someone for a miscalibrated emotion irrespective of that person's capacity to regulate the emotion away.

In Study 2.4, subjects read a vignette in which someone had a calibrated or miscalibrated emotion. Unlike in prior studies, we manipulated whether the individual was perceived as rational by stipulating that this person suffered an injury that affects his ability to think clearly and rationally. In a control condition, this person was severely injured but in a way that preserved his ability to reason. We conducted a pretest to ensure that (1) our manipulation successfully generated the inference that the target would be poor at emotion regulation and (2) that our control and experimental conditions were matched on overall injury severity.

Manipulation Pre-test. We recruited 97 subjects from Amazon's Mechanical Turk (mean age = 33.4, 43 reported female) to provide their impressions of a hypothetical friend named Jamie. Subjects were randomly assigned to read about Jamie either suffering a head trauma, which impairs his ability to think clearly and rationally, or hurting his back, which impairs his ability to move freely and easily. Full text is provided below. In all cases, subjects were asked to think of someone their age and sex. In the vignette the sex of Jamie matched the self-reported sex of the participant.

As expected, Jamie in the mental incapacity condition was judged to be less capable of controlling his emotions and mood (M = 3.38, SD = 1.53) compared to Jamie in the physical incapacity condition (M = 5.52, SD = 1.25), $t(89.05) = -7.52$, $p < 0.001$. Similarly, Jamie was rated as less capable of thinking clearly and rationally in the emotion incapacity condition (M = 3.62, SD = 1.60) compared to the physical incapacity condition (M = 5.82, SD = 1.12), $t(81.94) = -7.83$, $p < 0.001$. By comparison, when Jamie was described as suffering from a back injury, subjects judged him to be less physically capable (M = 4.08, SD = 1.34) than when he suffered a concussion (M = 4.83, SD = 1.20), $t(94.83) = 2.91$, $p = 0.004$. Despite these differences, subjects were equally sympathetic toward the mentally incapacitated Jamie (M = 6.06, SD = 1.21) and physically incapacitated Jamie (M = 5.86, SD = 0.99), $t(89.18) = 0.91$, $p = 0.367$. However, participants judged Jamie to be slightly less likable in the mental incapacitation condition (M = 5.57, SD = 1.43) relative to the physical incapacitation condition (M = 6.14, SD = 0.93), $t(78.18) = -2.30$, $p = 0.024$.

Our prediction was that subjects who read about mentally incapacitated Jamie would judge his ability to regulate away a miscalibrated emotion no differently than his ability to regulate away a calibrated emotion. This lack of an increase in perceived control in the miscalibrated condition should result in a corresponding lack of increased unsupportive behavior or decreased supportive behavior. By contrast, in the physical trauma condition, there is nothing stopping Jamie from reappraising his emotion. Therefore, we should observe the same increase in perceived emotion control (and corresponding unsupportive behavior). Critically, these two conditions were matched on how generally sympathetic Jamie is after the two injury types. Therefore, if, as we predict, we observe more supportive reactions, and fewer unsupportive reactions, toward Jamie in the mental incapacitation condition, this cannot be because participants felt a general desire to treat Jamie better due to his condition.

Study 2.4 Main Study Methods

Participants. We recruited 399 people from Amazon's Mechanical Turk (207 reported Female, mean Age = 37.7). This sample size yielded >95% power to detect an 50% attenuated interaction (or greater) based on the original effect size (*d* = .8) observed in Study 2.3.

Design and procedure. We used a crossed 2 (calibration: high vs low) x 2 (ailment: physical vs mental) between-subjects experimental design. Subjects were randomly assigned to one of the four experimental conditions at the beginning of the experiment.

The complete vignette was divided between two pages. On the first page, subjects read a short passage setting up the scenario and then revealing that Jamie had been in an accident that resulted in either mental or physical impairment. The full text is below:

You are about to pick your friend Jamie up from the airport. You are both headed to a mutual friend's wedding and you have agreed to carpool there together.

Mental Ailment Condition:

A few months ago, Jamie's car was struck from behind while he was waiting at a red light. The accident gave Jamie a concussion. While he has mostly recovered, he is still dealing with side effects from the head trauma. For instance, he has difficulty thinking clearly or rationally. He also has difficulty controlling his thoughts and moods. You've seen him struggle with this ever since the accident. Luckily, the doctors strongly believe Jamie will be fully recovered in about a month.

Physical Ailment Condition:

A few months ago, Jamie's car was struck from behind while he was waiting at a red light. The accident hurt his back. While he has mostly recovered, he is still dealing with side effects from the injury. For instance, he still has some stiffness and soreness in his back. He also has difficulty walking at a normal pace and getting up when he is sitting. You've seen him struggle with this ever since the accident. Luckily, the doctors strongly believe Jamie will be fully recovered in about a month.

On the next page, subjects answered two questions about the vignette, including what destination was stipulated in the vignette, and what ailment Jamie was suffering from. Subjects then received the rest of the vignette, which included the personal relevance manipulation, as well as our DVs, on the next page. The rest of the vignette read as follows:

> You pick Jamie up from the airport. But instead of looking excited about the trip he seems lethargic and in low spirits. This does not change as you drive to the wedding. When you bring up the wedding, he tries to show some enthusiasm but it is painfully obvious that his mind is elsewhere. He looks unhappy the entire time. In fact, at several points in the drive he seems to be on the verge of crying.
>
> It is clear that Jamie is upset about something. If he stays this way, people at the wedding will be able to tell and you are certain that it will detract from the happy day.
>
> **Calibrated emotion condition:**
>
> When you ask him what is going on, Jamie tells you that he was talking to his mother right before the trip started and she gave him some news. Apparently, she just learned that a friend of his from high-school named Tommy has committed suicide.
>
> **Miscalibrated emotion condition:**
>
> When you ask him what is going on, Jamie tells you that he was talking to his mother right before the trip started and she gave him some news. Apparently, the

National Institute of Health released a report saying that teen suicide increased by 1% last year.

When Jamie stops talking, he shrinks into his seat and resumes staring off into the distance. You have about half an hour left in your drive.

You know that Jamie has been having a rough time because of his recent car accident and the [mental | physical] problems it has caused him. But you are worried about Jamie's state being a drain on the festivities about to unfold.

How would you think about and react to Jamie in this situation?

Below the vignette were eight statements for which subjects indicated their agreement on a seven-point rating scale (1: strongly disagree; 7: strongly agree). One item measure perceived emotional control, "If Jamie wanted to, he could choose to stop feeling upset.". Two items measured perceived emotion calibration ("It objectively makes sense for Jamie to feel as upset as he does about what happened." and "Jamie's reaction to the news he just learned is appropriate."). Two items measured supportive reactions ("I would feel a great deal of sympathy for Jamie feeling this upset." and "I would do everything I could to accommodate Jamie."). Two items measured unsupportive reactions ("I would criticize Jamie for feeling this upset." and "I would express frustration toward Jamie for feeling this upset."). And one item measured perceived emotion strength ("Jamie is feeling extremely upset."). The eight items were shown in random order. On the following page subjects were asked to briefly recall what Jamie was upset about.

Subjects responded to a short demographics form which asked for their age and sex and were debriefed.

Results

We averaged our two supportive reaction (*r* = 0.68), unsupportive reaction (*r* = 0.79), and perceived emotion calibration (*r* = 0.84) items to create composite supportive reaction, unsupportive reaction, and emotion calibration ratings, respectively. Means and standard deviations for each of our five DVs are located in Table 2.2, below.

As planned, we conducted a series of ANOVAs regressing each our DVs on ailment type (mental vs physical), calibration (low vs high) and the interaction of ailment type and calibration. As expected, we observed a main effect of emotion type on our composite perceived calibration ratings, $F(1, 395) = 421.29$, $p < 0.001$, $\eta_G^2 = 0.52$, such that emotions in the low calibration condition were judged as less calibrated (M = 3.35, SD = 1.67) than high calibration condition (M = 6.18, SD = 1.01). Also as expected, there was no main effect of ailment type, $F(1, 395) = 0.03$, $p = 0.871$, or interaction, $F(1, 395) = 0.06$, $p = 0.80$. As in Study 2.3, we observed a small effect of calibration on perceived emotion strength. Subjects perceived Jamie's emotional reaction in the low calibration condition as slightly less severe (M = 5.97, SD = 1.07) than in the high calibration condition (M = 6.34, SD = 1.13), $F(1, 395) = 11.05$, $p = 0.001$, $\eta_G^2 = 0.03$.

Table 2.2
Means (and SD) for each of the five judgments in Study 2.4 across the four conditions.

	Mental Ailment		**Physical Ailment**	
DV	Low calibration	High calibration	Low calibration	High calibration
Perceived calibration	3.35 (1.66)	6.15 (1.03)	3.34 (1.68)	6.21 (1.00)
Perceived strength	5.91 (1.11)	6.33 (1.13)	6.03 (1.04)	6.35 (1.13)
Perceived control	2.66 (1.34)	2.50 (1.59)	3.76 (1.57)	2.70 (1.63)
Unsupportive reactions	2.03 (1.34)	1.82 (1.40)	2.46 (1.67)	1.66 (1.24)
Supportive reaction	5.46 (1.23)	6.21 (0.96)	5.07 (1.50)	6.13 (1.00)

Note: All ratings made on a 1-7 rating scale.

We next turned to our main DVs of interest: perceived control, supportive reactions, and unsupportive reactions. There was a significant effect of calibration on perceived control in the physical ailment condition such that subjects rated the low calibration reaction (M = 3.76, SD = 1.57) as more controllable than the high calibration emotion (M = 2.70, SD = 1.63), $t(196.15) = 4.65$, $p < 0.001$, 95% CI [0.61, 1.5], $d = .66$. However, there was no effect of calibration in the mental ailment condition, $t(196.96) = 0.81$, $p = 0.421$, 95% CI [-0.24, 0.58], $d = .11$. The combination of these resulted in the predicted interaction of ailment and calibration on perceived control, $F(1, 395) = 8.24$, $p = 0.004$, $\eta_G{}^2 = 0.02$. We therefore replicated the effect of calibration on control that we observed in prior studies and then completely attenuated it by stipulating that the target in question was temporarily incapable of reasoning clearly or rationally.

Also replicating prior studies, subjects were less supportive in the low calibration condition (M = 5.25, SD = 1.39) relative to the high calibration condition (M = 6.17, SD = 0.98), $F(1, 395) = 57.97$, $p < 0.001$, $\eta_G{}^2 = 0.13$. There was no effect of ailment type on supportive reactions, $F(1, 395) = 3.83$, $p = 0.051$, $\eta_G{}^2 = 0.01$. We also did not observe the predicted interaction of ailment type and calibration, $F(1, 395) = 1.60$, $p = 0.207$. In a planned moderated-mediation analysis, we replicated prior studies which found that perceived control partially mediated the effect of calibration on supportive reactions ($a = -0.61$, $p < .001$, $b = -0.28$, $p < .001$; $ab = 0.25$, 95% CI [0.07, 0.44]). However, contrary to our expectations, this overall effect was not moderated by ailment type *moderator*c* = 0.30, $t(395) = 1.26$, $p = .207$. As expected, ailment type did not moderate the effect of control on supportive reactions, *moderator*b* = 0.07, $t(393) = 0.95$, $p = .341$. See Figure 2.4 below for full model output.

Turning to unsupportive reactions, we replicated the effect of calibration on punishment such that subjects were more likely to be unsupportive toward Jamie in the low calibration condition (M = 2.26, SD = 1.53) than the high calibration condition (M = 1.75, SD = 1.33), $F(1, 395) = 12.59$, $p < 0.001$, $\eta_G^2 = 0.03$. However, this main effect of calibration was driven by a significant interaction of ailment type and calibration, $F(1, 395) = 4.28$, $p = 0.039$, $\eta_G^2 = 0.01$. Subjects were more unsupportive in the low calibration emotion (M = 2.46, SD = 1.67) relative to the high calibration emotion (M = 1.66, SD = 1.24) in the physical ailment condition, $t(187.87) = 3.87$, $p < 0.001$, 95% CI [0.39, 1.21], $d = .54$, but this effect was completely attenuated in the mental ailment condition, $t(194.82) = 1.08$, $p = 0.281$, 95% CI [-0.17, 0.59], $d = .15$. In a planned moderated-mediation analysis we found that perceived emotion control completely mediated the effect of emotion calibration on unsupportive reactions ($a = -0.61$, $p < .001$, b = 0.53, $p < .001$; $ab = -0.47$, 95% CI [-0.81; -0.15]). However, both the effect of calibration on control, *moderator*a* = 0.90, $t(395) = 2.87$, $p = .004$, and calibration on unsupportive reactions, *moderator*c* = -0.60, $t(395) = 2.07$, $p = .039$, were moderated by ailment type (see Figure 2.4, below). As expected, ailment type did not moderate the effect of control on punishment, *moderator*b* = -0.09, *t*(393) = 1.11, *p* = .268.

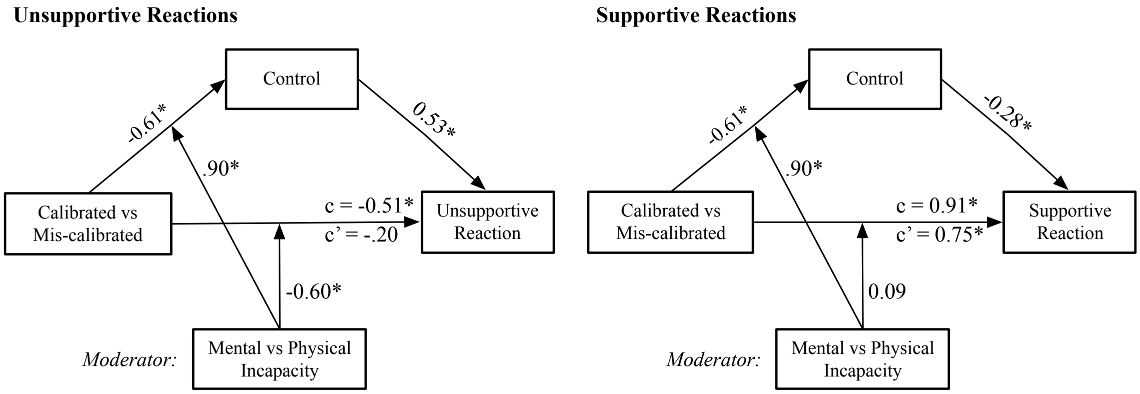

Figure 2.4. Moderated-mediation analyses for punishment and sympathy behaviors in Study 2.4. Control mediated the difference in punish and sympathy behaviors across low vs high personal relevance. This effect was moderated by the incapacity manipulation. Punishment, but not sympathy, was moderated by mental and physical capacity as well.

Discussion

Results from Study 2.4 supported the RER hypothesis. We replicated support for Hypotheses 1 and 2: perceived emotion control predicted supportive and unsupportive behaviors toward someone feeling a negative emotion. We also replicated findings from Studies 2.2-2.3 showing that lay people view miscalibrated emotions as more controllable than calibrated emotions, and that these changes in control predict changes in supportive and unsupportive behaviors. In this study, supportive behavior was partially mediated by perceived control while unsupportive behavior was fully mediated by perceived control.

Thus, we replicated findings showing support for the hypotheses H1-H4 we proposed in the Introduction and established support for H5.

Study 2.4 also showed that people's inferences about perceived emotion control in the face of miscalibrated emotions depends on their prior assessment that this person is generally has rational control over his thoughts. When we stipulated that the target could not think clearly because he had suffered a concussion, people no longer judged that person to have more control over the emotion in the miscalibrated condition. This had important downstream consequences for people's punishing behavior. Corresponding with a lack of control, people no longer exhibited unsupportive behavior, such as criticizing or expressing frustration, toward the target for his emotional reaction in the miscalibrated emotion condition (relative to the calibrated condition). This was not due to the mere fact that Jamie, the target, was injured in this condition. In the control condition, in which Jamie had suffered a severe back injury (and was pre-tested as being equally sympathetic and indeed more likable), subjects judged that he had more control over the miscalibrated emotion and were unsupportive toward him for it.

This finding also rules out an alternative explanation for subjects' unsupportive behavior in prior studies. Subjects judged the miscalibrated emotion as equally miscalibrated in both the mental and physical incapacitation conditions. However, subjects were only unsupportive toward Jamie for his miscalibrated emotion in the physical ailment condition, where he had the ability to cognitively regulate it. Thus, at least for unsupportive behavior, it appears that people are reasoning more about control than about the inappropriateness of the emotion per se. This is consistent with the RER, which stipulates that people act in an unsupportive manner as a mechanism for motivating them to regulate their emotion: doing so is only rational when the target has the capacity to follow through. However, in contrast to unsupportive behavior, we found that supportive behavior was less influenced by the target's emotion regulation capacity and control. As in prior studies, supportive behavior was partially mediated by perceived control; however, we still observed significance differences in supportive behavior in the mental incapacitation condition. It appears that supportive reactions, like feeling sympathy for someone, is heavily influenced by the normative status of someone's emotion in a way that unsupportive behaviors are not.

One limitation of Studies 2.1-2.4 is that they all use hypothetical vignettes which involve hypothetical friends. Thus, it is possible that our findings do not generalize to important, real-life behavior. We addressed this limitation in Studies 2.5 and 2.6. In Study 2.5 we conducted an autobiographical recall study to test whether perceived emotion control in response to actual close-other's emotions predicted supportive and unsupportive responses. In Study 2.6 we test whether perceived emotion control predicts people's policy attitudes in a current national debate regarding the responsibility that Universities have (or not) toward protecting minority students against micro-aggressions.

Study 2.5

Below we report an autobiographical recall task where we probed people's recent supportive and unsupportive behavior towards others. We asked subjects to write about a recent time that someone they knew felt a strong negative emotion and to report what they thought of the person at the time as well as how they behaved. Consistent with the

RER, we hypothesized that people’s supportive and unsupportive behavior towards their close other would vary with the degree of control they had attributed to the individual.

Methods

Participants. We recruited 298 subjects from Amazon’s Mechanical Turk (mean age 35.6, 143 reported Female). This yielded greater than 95% power to detect associations of $r = .20$ or greater.

Procedures. Subjects were told that we were conducting a study on how people behave in close relationships. We asked subjects to think of someone they are close to and with whom they frequently interact, such as a good friend, romantic partner, or family member. Subjects provided the (i) initials of the individual, (ii) the sex of the target, (iii) the person’s age, and (iv) what his or her relationship to the target was. We then instructed participants to try and think about a recent time that this person was around them while experiencing a strong negative emotion.

Please think of the most recent time that [initials] felt a strong negative emotion such as feeling sad or upset, anxious, or stressed in your presence. Try to think of a time when you had to react to this person feeling sad, upset, anxious, or stressed. This can be a case in which you acted in a supportive or unsupportive manner. When you have thought of a specific time, please press the arrow below to continue.

Subjects were asked what emotion best described the event they had thought of (from the list of “upset,” “anxious,” “stressed,” or “sad”) and how long ago the event took place. Subjects then reported what caused the emotion by completing the statement. For instance, if subjects had indicated that the close other felt “upset,” then they were instructed to complete the sentence fragment, "[Initials] felt [upset] about…". For this question, as well as all remaining ones, the initials and emotion type were dynamically inserted to match what subjects had reported at the start of the survey. Subjects answered remaining questions about the situation in two batches, reported below. For each set of questions, the original description of the emotion provided by the subject was shown at the top of the screen. All ratings were made on a 7-point scale (1 = *not at all accurate*; 7 = *completely accurate*).

Judgments about the emotion. On the next screen, subjects rated the accuracy of five statements about what they thought about the emotion in that situation. This included perceived (i) emotion strength (e.g., "At the time, I thought that [Initials] felt extremely upset"), (ii) how unfortunate it was the close other felt the emotion "At the time, I thought that it was unfortunate for [him] that [initials] felt [upset]", and (iii) emotion control "At the time, I thought that [initials] could choose to stop feeling [upset] if [he] tried hard enough." Subjects also responded to two statements about whether the emotion was diagnostic of something bad about the target, including (iv) "At the time, I thought that it reflected poorly on [Initials]'s moral character." and (v) "[Initials] felt [upset] because of something [he] was at fault for.".

Emotional and Behavioral Reactions. Finally, subjects reported how they felt and behaved in response to the target's emotion. Subjects were told that they would see a series of statements describing certain behaviors or feelings and that they should indicate for each one how accurately that statement described their reaction on a seven-point rating scale (1 = *not at all accurate*; 7 = *completely accurate*). Two items measured supportive thoughts including (i) "I felt sympathy for [initials] for feeling [upset]." and

(ii) "I felt bad for [initials] that [he] felt this [upset].". Two items measured unsupportive thoughts (iii) "I felt annoyed at [initials] for feeling this [upset]." and (iv) "I did not want to be around [initials] at this time.". Two items next measured supportive behaviors including (v) "I accommodated [initials] as much as [he] wanted." and (vi) "I spent a lot of effort and time trying to make [initials] feel better." Finally, two items measured unsupportive behaviors including (vii) "I criticized [initials] for feeling as [upset] as [he] did." and (viii) "I expressed frustration at [initials] for feeling this [upset].".

After doing the recall task, subjects completed a demographics form, which included information about their sex, age, political orientation, and religiosity.

Results

Subjects predominately recalled situations in which their spouse or romantic partner (*n* = 128), or close friend (*n* = 78) felt a negative emotion, typically feeling upset (*n* = 115) or sad (*n* = 103). On average, these were cases in which subjects judged that the other person's emotion was strong (*M* = 5.81, *SD* = 1.25) and that it was unfortunate for this person that they felt this way (*M* = 5.40, *SD* = 1.65). Additionally, on average subjects reacted in a supportive way: they largely accommodated their close other (*M* = 5.49, *SD* = 1.55) and put in time and effort to make him/her/them feel better (*M* = 5.29, *SD* = 1.55), while typically not criticizing (*M* = 1.96, *SD* = 1.59) or expressing frustration toward that individual (*M* = 2.34, *SD* = 1.85). In the remaining analyses, however, we examined what judgments about the emotion correlated with decisions to judge or react in a supportive or unsupportive manner. As planned, we created composite measure of supportive thoughts (*r* = 0.69), unsupportive thoughts (*r* = 0.72), supportive behaviors (*r*

= 0.66), and unsupportive behaviors ($r = 0.69$) by averaging together subjects' accuracy ratings for the two statements for each construct. Table 2.3 below contains summary statistics and correlations from our primary dependent measures of interest.

Table 2.3
Descriptive statistics and correlations between primary measures of interest in Study 2.5.

	Dependent Variable	**Mean (SD)**	**1**	**2**	**3**	**4**	**5**	**6**
Attribution	1. Control	2.97 (1.84)						
	2. Fault	2.63 (1.93)	0.30					
	3. Moral Character	2.07 (1.69)	0.51	0.48				
Reactions	4. Supportive Thoughts	5.70 (1.38)	-0.41	-0.19	-0.37			
	5. Unsupportive Thoughts	2.39 (1.70)	0.47	0.41	0.61	-0.55		
	6. Supportive Behavior	5.39 (1.41)	-0.27	-0.15	-0.24	0.61	-0.52	
	7. Unsupportive Behavior	2.15 (1.58)	0.49	0.46	0.67	-0.51	0.80	-0.45

Notes. All ratings were made on a 1-7 rating scale.
df = 296
All correlations significant at $p < 0.001$ except for r(fault, supportive behavior) which was significant at $p = 0.01$.

As planned, we examined the correlation between these four reactions and subjects' attributions of emotion control, fault over the situation, and moral character. Consistent with past work, perceptions of fault and moral character were associated with all four reaction types (ps <= 0.01). However, consistent with the RER hypothesis, situations in which subjects tended to attribute greater emotion control to their close other were also situations in which they were less likely to have supportive thoughts of that person, $r(296) = -0.41$, $p < 0.001$, 95% CI [-0.50, -0.31], less likely to act in a supportive manner, $r(296) = -0.27$, $p < 0.001$, 95% CI [-0.38, -0.17], as well as more likely to have unsupportive thoughts towards that person, $r(296) = 0.47$, $p < 0.001$, 95% CI [0.37, 0.55], and more likely to behave in an unsupportive manner, $r(296) = 0.49$, $p < 0.001$, 95% CI [0.40, 0.57] (see Figure 2.5, below). We next tested whether control predicted supportive and unsupportive reactions after accounting for the shared variation in perceived fault and moral failing. Consistent with the RER, we found that an increase in control predicted a decrease in supportive thoughts (*b* = -0.22, *SE* = 0.05, *t* = -4.97, *p* < 0.001) and behavior (*b* = -0.15, *SE* = 0.05, *t* = -3.1, *p* = 0.002), and an increase in unsupportive thoughts (*b* = 0.19, *SE* = 0.05, *t* = 3.89, *p* < 0.001) and behavior (*b* = 0.16, *SE* = 0.04, *t* = 3.87, *p* < 0.001), even when accounting for variation in fault and moral character judgments.

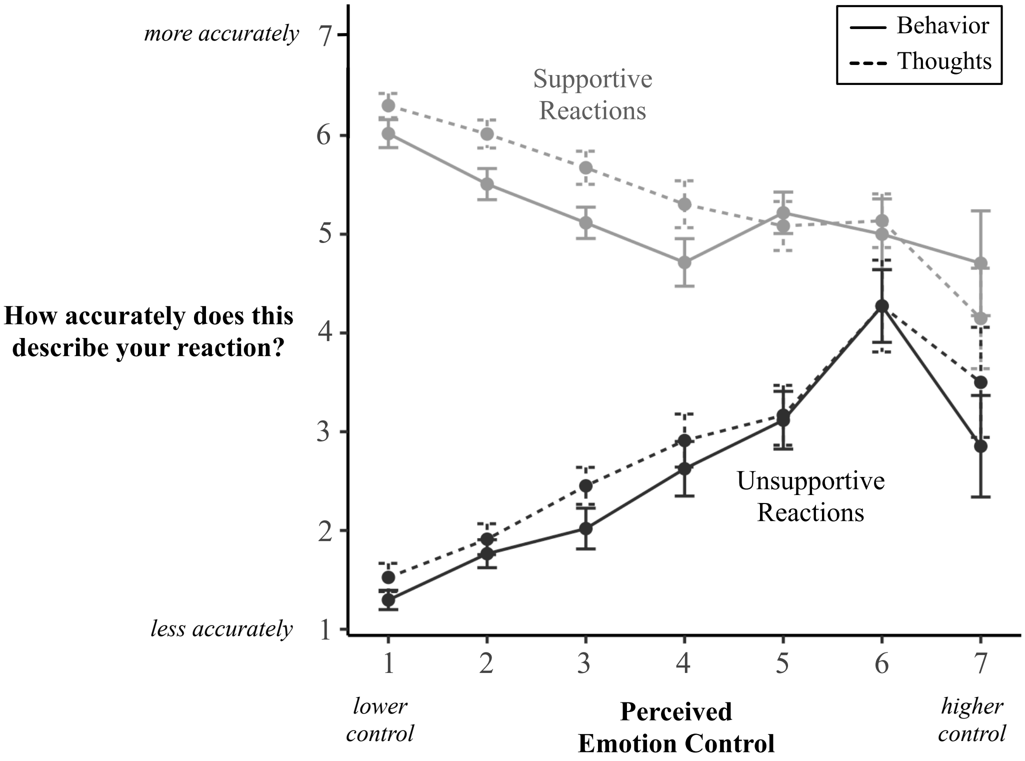

Figure 2.5. Means and standard error for agreement that statements describing supportive and unsupportive thoughts and behavior accurately described past behavior toward a close other. Accuracy judgments are grouped by perceived control over the emotion at the time.

We noticed that most subjects (53%) recalled situations in which they judged the emotion as neither diagnostic of immoral character (moral character ≤ 2) nor the result of something that was that individual’s fault (fault ≤ 2). If the RER hypothesis is correct, then we should observe that control predicts people's supportive and unsupportive reactions towards others even in these cases. We conducted a series of exploratory tests measuring the strength of association between control and subjects' self-reported reactions. Consistent with the RER hypothesis, control correlated with supportive thoughts, $r(155) = -0.30$, $p < 0.001$, 95% CI [-0.44, -0.15], unsupportive thoughts, $r(155) = 0.35$, $p < 0.001$, 95% CI [0.20, 0.48], and unsupportive behaviors, $r(155) = 0.23$, $p < 0.001$, 95% CI [0.08, 0.37], but not supportive behaviors, $r(155) = -0.15$, $p = 0.06$, 95% CI [-0.30, 0.01][4]. Thus, control predicts people's supportive and unsupportive behavior in real-life situations absent people's considerations of how moral deservingness.

Discussion

Study 2.5 investigated whether variation in everyday cases of supportive and unsupportive behavior is associated with people's judgments that the person they are reacting to has control over the emotion in question. Consistent with the RER hypothesis, we found that control predicted sympathetic thoughts (such as feeling sympathy), sympathetic behavior (such as putting effort into helping the person feel better), unsympathetic thoughts (like feeling annoyed at the individual), and unsympathetic

[4] When we conducted a similar exploratory exercise investigating trials in which the subject reported that the target's emotion was not controllable (control ≤ 2) or their ultimately fault (fault ≤ 2), we found that perceived moral character did not predict any of the four reaction types ($n = 111$, rs = –0.06 – 0.18, ps ≥ 0.06). Similarly, perceived fault did not predict any of the four reaction types in the set of trials which were uncontrollable and not the person's fault ($n = 138$, rs = –0.08 – 0.16, ps ≥ 0.06). It would be wrong to conclude from this that perceived fault or immorality play no role in people's decisions to be supportive or unsupportive. Rather, we should conclude that perceived emotion control is a particularly robust predictor of support in everyday interpersonal interactions whereas fault and morality are not.

behavior (like expressing frustration toward them). Critically, we found that control predicted these reactions independently from other canonical predictors of sympathy and punishment behavior; namely, perceived causal fault and perceptions of poor moral character. We established this in two ways. First, we found that variation in perceived control predicted supportive and unsupportive reactions when accounting for shared variation in fault and character judgments. Second, we found that control predicted people's reactions even when restricting our analysis to situations in which subjects explicitly indicated that the person was not at fault for whatever caused their emotion, nor was being judged as having poor character. This study strongly suggests that the results we found in subjects' reactions to hypothetical vignettes in Studies 2.1-2.4 replicate and predict behavior in real-life, close interpersonal relationships.

We were next interested in whether the RER could explain people's attitudes towards others' emotions outside the context of close interpersonal relationships. Organizations spend a great deal of resources minimizing the emotional suffering of others. To name one example, the Make A Wish foundation raises millions of dollars a year with the mission of addressing the emotional suffering that children and parents face when dealing with terminal illnesses. They write, "For wish kids, just the act of making their wish come true can give them the courage to comply with their medical treatments" and "Parents might finally feel like they can be optimistic". We hypothesized that when assessing the value of these kinds of missions, people reason about how much control the target beneficiaries have over their emotions and that these attributions affect their support for policies or other initiatives aimed at helping reducing emotion suffering. To test this hypothesis, we investigated one topic that has recently gain a great deal of attention in the United States: University policies aimed at protecting minority students from microaggressions.

Study 2.6

In Study 2.6 we tested whether the RER hypothesis could help explain lay people's reasoning in debates about whether universities ought to protect minority students from others' microaggressions against them. Some scholars have argued that anti-microaggression initiatives, such as trigger warnings, safe spaces, and bans on the use of black face (or other offensive behavior), come with high costs. These costs include creating a culture of political correctness or impinging on people's freedom of expression, thereby violating the spirit of open inquiry and authority-independent intellectual pursuit (Lukinoff & Haidt, 2019). Of course, anti-microaggression initiatives may be worth these costs if they are necessary for improving the mental well-being of minority students. One source of opposition to these initiatives may stem from judgments that these initiatives are, in fact, not necessary for improving minority students' mental health.

If people believe minorities are able to regulate their emotional reactions to microaggressions, thereby protecting or improving their mental health on their own, then they may think that policies aiming to protect minorities are unnecessary. Indeed, many commentators on debates surrounding microaggressions have argued that, either because most micro-aggressive behavior is highly ambiguous with respect to racial animus, or because the violation in question is minor, that strong emotional reactions to them are miscalibrated (e.g., Lilienfeld, 2015). Instead of suffering from legitimate grievances, these scholars argue, minorities who complain about microaggressions are "hypersensitive" to these perceived slights and injustices. As we observed in Studies 2.2-2.4, judgments of emotion calibration affect attributions of control, which then affects supportive and unsupportive reactions. Consistent with this reasoning, we hypothesized that people who judge that others have a high degree of personal control over their emotion, especially in the face of microaggressions, will be less supportive of anti-microaggression initiatives.

Methods

Participants. 299 people (152 reported female, mean age = 36.6) were recruited from Amazon's Mechanical Turk. This yielded greater than 95% power to detect associations of $r = .20$ or greater.

Procedure. In the first part of the study, subjects read short vignettes about three minority students, an African American student named Shane, a student of Asian descent named Amy, and a student of Latin descent named Manuel. Each vignette mentioned that the student in question was exposed to microaggressions, was upset or anxious because of them, and was either doing poorly in classes or considering changing schools because of them. Across the three vignettes, subjects read about a variety of different microaggressions including, Shane: *crossing the street when the student approaches*, *staring at the student*, *wearing black face*; Amy, *asking about her perspective as minority, mistaking her for another person of similar ethnicity*; Manuel, *lowering* *expectations relative to other students, mistaking for someone of lower status (e.g., janitor)*[5]. For instance, in the "Shane" scenario, subjects read:

> Shane is an African American political science major at a high-ranking public university in the American Midwest. Shane recently told his academic advisor that his college experience has been upsetting on account of his status as an African American, that he feels depressed, and that he is thinking of transferring schools. When asked for examples, Shane responds, "Other students cross the street when they see me walking their direction, or they stare at me whenever I'm around them. During Halloween, I see students dress up in black face. Even though I work hard, this stuff makes me feel like I don't belong here."

Subjects provided six judgments in response to each vignette. Of primary interest were two items measuring perceived emotion control: (1) "Even if these events initially upset Shane, he can choose to stop being upset by them if he tries." (2) "Shane has control over how upset he feels about other people's behavior." Two items measured how rare the subject though microaggressions were, including (3) "Shane's experience is probably rare (even amongst other minority students at his school)." and (4) "The type of behavior that Shane describes is uncommon at his University." And finally, two items measured general impressions and likability of the student: (5) "Shane does not seem like a likable person." and (6) "Shane does not seem like someone easy to get along with." Subjects rated their agreement with each of these six statements on a seven-point rating scale (1: do not agree at all, 7: completely agree). Each of the six statements was

[5] Microaggressions were drawn from the Racial and Ethnic Microaggressions Scale (Nadal, 2011). See Appendix I for full text of each of the three vignettes.

presented in a random order, and each of the three vignettes was shown in a random order.

Subjects were then asked to give their attitudes about anti-microaggression initiatives. To reduce competition-neglect, subjects were told the following:

> University administrators at these institutions are aware of the experiences of students like Shane, Amy, and Manuel. In response, these administrators are considering enacting policies that they think will improve the experience for minority students. Specifically, they are considering investing money and space to create "safe spaces" where minority students can spend time and share their experiences. They are also considering enacting rules banning certain student activities that are offensive to minorities (like wearing black face) and having faculty members and graduate students attend sensitivity training.
>
> Even if these policies are likely to help, the administrators at these institutions share a few major concerns about them. First, they require significant investment of limited university resources. Second, they involve limiting the freedom of faculty and non-minority students. And third, they may create a culture of political correctness that negatively impacts people's ability to have open and frank discussions about important issues.

Subjects were then asked to indicate their attitude toward four different anti-microaggression initiatives by rating their agreement (1: strongly disagree, 7: strongly agree) with four statements. These included, (1) "Universities should spend time and money to create 'safe spaces' for minority students," (2) "Universities should ban certain activities if minority students find those activities offensive," (3) "Universities should mandate that faculty undergo sensitivity training to reduce the occurrence of micro aggressions," and (4) "Universities should regularly notify new students about what behaviors are considered rude or offensive." These four statements were shown in a random order.

After this, subjects filled out a demographic questionnaire which measured their age, sex, ethnicity, political orientation, religiosity, and religious orientation.

Results

As planned, we created composite ratings of perceived emotion control ($r = 0.79$), microaggression rarity ($r = 0.76$), and likability ($r = 0.81$), by averaging together subjects' agreement with each of the two control, rarity, and likability items. Then, for each subject, we averaged their control ($\alpha = 0.95$), rarity ($\alpha = 0.91$), and likability ($\alpha = 0.92$) ratings across the three vignettes. We then created a composite policy support measure by averaging together subjects' agreement with the four policy statements ($\alpha = 0.84$). The analyses we report below are based on these four composite variables.

As predicted, the greater control that subjects attributed to minorities over their emotional reactions to others' behavior, the less supportive they were of university policies to $r(297) = -0.30$, $p < 0.001$, 95% CI [-0.40, -0.19] (Figure 2.6). However, control was also associated with subjects' other evaluations as well. The more rare subjects thought microaggressions were, the more control they attributed to minorities over their reaction, $r(297) = 0.43$, $p < 0.001$, 95% CI [0.33, 0.51]. The greater emotion control was associated with finding the minority students less likable, $r(297) = 0.34$, $p < 0.001$, 95% CI [0.23, 0.43]. Rarity and likability were also strongly associated with policy support in expected ways. The more rare subjects thought microaggressions were, the less supportive they were of policy to address them, $r(297) = -0.25$, $p < 0.001$, 95% CI: [-0.35, -0.14]. And subjects who found minorities to be less likable also reported less support for policy, $r(297) = -0.25$, $p < 0.001$, 95% CI [-0.35, -0.14].

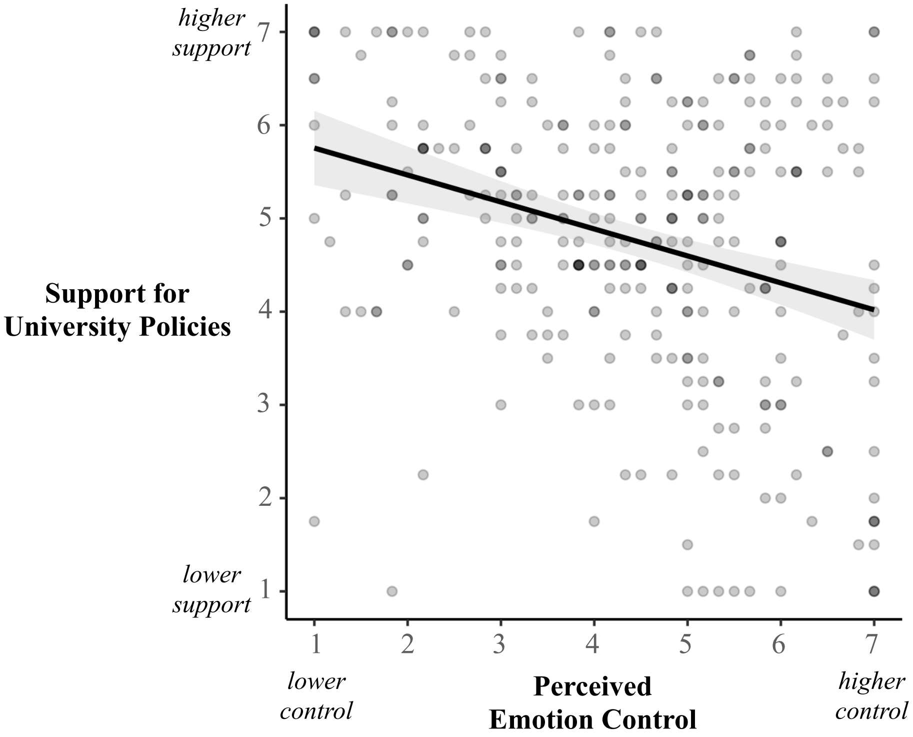

Figure 2.6. The association between subject's average attribution of emotion control to minorities and their stated support for anti-microaggression university initiatives in Study 2.6. The more control subjects attributed to minorities over their emotions, the less supportive they were of university policies to prevent microaggressions.

We next tested whether control predicted policy support when accounting for the variance explained by perceived rarity and liking. When regressing policy support on control, liking, and rarity, we found that control was still a significant predictor of policy attitudes, *b* = -0.19, *SE* = 0.06, *t* = -3.31, *p* = 0.001. Rarity, too, was still a significant predictor of policy support, *b* = -0.19, *SE* = 0.08, *t* = -2.45, *p* = 0.015. However, subjects liking of the minorities in the vignettes no longer predicted policy support, *b* = -0.07, *SE* = 0.07, *t* = -0.88, *p* = 0.378. Thus, considerations about how necessary the policies are, including how common the behavior they are trying to prevent is, as well as how much the population they are protecting requires their help, appear to be more robust predictors of policy support than general impressions of the individuals in question.

Finally, we conducted an exploratory analysis testing how political orientation predicted each of our variables. More conservative subjects liked the minority students less, *r*(294) = 0.34, *p* < 0.001, 95% CI [0.24, 0.44], believed that microaggressions were rarer, *r*(294) = 0.39, *p* < 0.001, 95% CI [0.28, 0.48], and attributed more control to the minorities over their emotional reactions, *r*(294) = 0.35, *p* < 0.001, 95% CI [0.25, 0.45]. Unsurprisingly, then, conservatives were also less supportive of anti-microaggression policy, *r*(294) = -0.33, *p* < 0.001, 95% CI [-0.42, -0.22]. However, our exploratory analysis also revealed that control was still a significant predictor of policy support even when accounting for political orientation (as well as rarity and liking), *b* = -0.15, *SE* = 0.06, *t* = -2.50, *p* = 0.013. Thus, variation in perceived emotion control amongst liberals and conservatives explains variation in supportive for anti-microaggression policies.

Discussion

Study 2.6 tested whether predictions made by the RER hypothesis extend outside the domain of close relationships. We predicted that people who judged emotional reactions in response to microaggressions to be highly controllable would also judge costly policies to be less necessary for helping those students, and so would oppose them. This prediction was confirmed: even when controlling for general impressions of minority students, the rarity of micro-aggressions, and subjects' political orientation, we found that differences in the perceived controllability of emotions predicted attitudes towards initiatives designed to prevent microaggressions from occurring.

General Discussion

When someone feels anxious, embarrassed, upset, or distressed, they benefit from seeking out and receiving sympathy from others. But it is not guaranteed that observers will help them. Moreover, observers sometimes intentionally make the sufferer feel even worse for feeling the negative emotion. Past work has shown that observers' antisocial reactions to others' suffering can occur when people judge it as especially costly to help (e.g., Cameron et al., 2019) or when the sufferer is judged as deserving their pain (or at least undeserving of sympathy; e.g., Haslam, 2006; Weiner, 1995). However, it is unlikely that considerations of burden or moral deservingness are the whole story. Observers often refuse to help, and are deliberately unsupportive towards, sufferers who feel bad for things that are not their fault and which are not diagnostic of poor character. What explains people's behavior in these situations?

We hypothesized that observer's supportive and unsupportive reactions reflect a tendency to monitor and motivate sufferers' attempts to improve their emotional well-being on their own. We call this the *regulating emotion regulation* hypothesis (RER). According to the RER, when people are exposed to someone feeling an undesirable emotion, they reason about whether that person has the ability to regulate away that emotion themselves. If the sufferer cannot, then they are in genuine need of help and the observer (as long as she has sufficient regard for this person's plight) will take on the burden of accommodating and sympathizing with them. However, if the sufferer is judged as capable of regulating away their negative emotion themselves, then the observer will try to reduce their negative emotion indirectly by motivating the sufferer to engage in intrapersonal emotion regulation. This latter behavior sometimes entails blaming, punishing, or otherwise acting in an unsupportive manner toward the sufferer. Ironically, this entails making someone who feels bad feel even worse in order to eventually make them feel better.

One prediction of the RER hypothesis is that observers' supportive and unsupportive reactions are based on their assessment of the sufferer's control over his or her emotion. Consistent with this prediction, we observed across all our experiments that perceived emotion control (operationalized by judgments that someone could stop feeling bad if they wanted to) predicted observers' supportive and unsupportive behavior. Study 2.1 found that individual differences in perceived emotion control correlated with sympathy and agreement that subjects would try to make the sufferer feel bad for their emotion. In Studies 2.3-2.4, we manipulated how much control subjects attributed to an emotion target. Across these studies, subjects who were randomly assigned to the low emotion control condition were more supportive and were less unsupportive compared to those assigned to the high-emotion control condition. In Study 2.5 we found that people's tendency to feel supportive or unsupportive thoughts, and to behave in supportive or unsupportive way towards close others in recalled real-life situations was robustly predicted by their attributions of emotion control. And finally, in Study 2.6, people who tended to view the victims of microaggressions as capable of regulating their negative reactions away were less supportive of policies that would intervene to prevent microaggressions in the first place. These findings are consistent with past work showing that people's helping and punishing behavior is heavily influenced by attributions of so-called "offset control" – someone's control over improving their own situation (Brickman et al., 1982; Karasawa, 1991; Meyer & Mulherin, 1980). Viewed in this light, our findings show that this reasoning extends to how people reason and react to emotional suffering.

If supportive and unsupportive behavior is sensitive to attributions of emotion control, this raises the question how people arrive at their attributions of emotion control in the first place. In Study 2.1, we hypothesized that one cue would be people's internal, dispositional expectations about emotion control. Past work has shown that people hold different implicit theories of emotion as being either controllable or uncontrollable (Tamir et al., 2007). These theories are influenced in part by their own history of successfully regulating their emotions and predict a greater motivation and tendency to reappraise emotions in the future, as well as to have more successful outcomes regulating emotion (e.g., DeCastella et al., 2013; Ford et al,. 2018). We found that, indeed, people apply their implicit theory of emotion control to others: people who generally believe they themselves have control over emotions attributed control to others as well. These individuals were then indirectly also less likely to feel sympathy for others, and more likely to react in an unsupportive manner.

However, no work to date has established a causal relationship between attributions of emotion control and supportive or unsupportive behavior. Moreover, no work has studied how people attribute emotion control to others across situations. We addressed this gap by speculating that people would generally be sensitive to features of an emotion or situation that makes intrapersonal regulation strategies more or less successful. Based on this, we hypothesized that people would judge miscalibrated emotions to be more controllable than calibrated emotions. This hypothesis was supported in Studies 2.2–2.4. In these studies, we kept the description of how severe someone's emotional reaction was constant across conditions, while varying what it was that they were upset, embarrassed, or distressed about. In "low calibration" conditions, the trigger was something of low personal relevance (e.g., a statistic about suicide prevalence) or low severity (e.g., a pile of sudoku books being destroyed). Study 2.2 showed that in these conditions, people judged the emotion as less calibrated to the situation and therefore more controllable. Studies 2.3 and 2.4 built on this to show that supportive and unsupportive reactions to miscalibrated emotions were mediated by the aforementioned differences in control. Most importantly, Study 2.4 showed that this effect was moderated by prior expectations that the individual in question is capable of thinking rationally and clearly – that is, that she has the capacity to use information about the objective circumstances to reappraise her emotion.

We have argued that an emotion's calibration to a situation is a cue of its controllability; however, an emotion's calibration is a feature that may affect people's behavior in other ways. For instance, miscalibrated emotions are often judged as deviant, and the people who experience them as immoral, irrational, or hyper-sensitive (e.g., Gromet et al., 2016; Szczurek, et al, 2012). It is also possible that people react negatively to others miscalibrated emotions directly – simply preferring not to show support for inappropriate emotional reactions irrespective of their general impressions of the other person or attributions of emotion control. We addressed these concerns in Study 2.4. When it was stipulated that the sufferer could not regulate his emotion away, in this case because he had suffered a concussion that interfered with his ability to think rationally and clearly, observers no longer judged him more capable of regulating away a miscalibrated emotion relative to calibrated one. Critically, despite the fact that people still judged the sufferer as having a miscalibrated emotion, they withheld criticizing and expressing frustration toward him[6]. Thus, miscalibrated emotions on their own are not enough to result in unsupportive reaction. Rather, consistent with the RER hypothesis, attributions of control appear to be necessary.

Although most of our analysis has concerned how people react towards others in close relationships, the RER hypothesis may help shed light on recent debates about what responsibility third parties, like universities, have to helping others deal with trauma and emotionally charged situations. In Study 2.6 we test this possibility in the case of microaggressions. We found that people who think that microaggressions are easy to emotionally recover from are less supportive of costly policies that intervene to help students. These findings suggest that advocates of anti-microaggression policies may benefit from providing opponents information relevant to the ease or difficulty of

[6] This finding supports our claim that people typically use calibration as a cue of potential emotion control, but it does not entail that people never punish others for having inappropriate or strange emotions. After all, some work has shown that people endorse the death penalty for criminals who experience schadenfreude while committing heinous crimes (e.g., Gromet, et al, 2016). We can reconcile these findings by noting that, in everyday contexts, miscalibrated emotions are not severe enough to be treated as diagnostic of poor dispositions or character. Instead, they are diagnostic of a temporary and correctable lack of cognitive effort or clear thinking.

emotion regulation. For instance, some work has shown that repeated exposure to micro-aggressions is associated with poorer mental health and increased feelings of alienation (West, 2019). By contrast, opponents may benefit from arguing that policies have a net negative on minority mental health, by making emotion regulation later in life even more difficult (see Lukinoff & Haidt, 2019, for such an argument). In sum, our findings suggest that it will be helpful to keep in mind lay considerations of who is viewed as responsible for maintaining emotional health, which will be determined by judgments of who is viewed capable of doing so.

The rationale (and rationality) of regulating others' emotion regulation

Our most striking finding is that, in Studies 2.1, 2.3, and 2.4, people reported a desire to make a close other feel *even worse* when he or she is already suffering. And, in Study 2.5, we found that people reported past behavior in which they judged, and purposely treated, a close other in unsupportive ways. While on its face counter-intuitive, we argued that this behavior reflects a goal to motivate that person to regulate their emotion away themselves. Support for this comes from the observation that the behaviors subject reported (i.e., making someone feel bad, expressing frustration, and criticizing) are strongly associated with blaming behaviors that people conduct in order change or modify others' conduct (Study 2.3 pretest). Second, people only report these behaviors when it would be rational to do so in light of that goal – i.e., when the target had control over the emotion and so could constructively respond to the motivation. In light of this finding, a valuable goal for future research is to investigate whether people's decision to behave in an unsupportive manner ultimately succeeds in improving someone's emotional well-being. That is, is pressuring and criticizing others an effective strategy for ultimately improving their emotional state?

Some work suggests that motivating people to regulate their emotions is likely to succeed and ultimately be beneficial for them. People who believe that emotions are highly controllable tend to be more motivated to regulate them, and more successful at regulating them (e.g., De Castella et al., 2013; Ford et al., 2017; Schroder et al., 2015; Tamir et al., 2007). Additionally, past research has shown that it is possible to train people to better regulate their emotions (e.g., Finkel, Slotter, Luchies, Walton, & Gross, 2013; see Cohen & Oschner, 2018, for a recent review). From this we may predict that communicating to others that their emotions are under their control and motivating them to exert said control over them, will help those individuals recover from trauma.

However, there is an important difference between teaching someone how to regulate their emotions – what clinicians do – and merely telling that person that she should be able to feel better and pressuring her to do so – what we have documented here that lay people do (Study 2.5). Indeed, recent experimental work suggests that people do not cope with negative emotions better when simply told that they should be able to. Kneeland et al (2016a) had subjects write paragraphs defending the idea that emotions are controllable or, in another condition, uncontrollable. When subjects were later induced to feel sad, those that were induced to believe that emotions were controllable engaged in more strategies to try and regulate their emotional experience. Critically, however, these subjects did not actually feel better relative to those manipulated into thinking they were uncontrollable. And, in a related study, subjects who were manipulated to believe that emotions are not controllable were more accepting of their emotions and blamed themselves less for experiencing them, resulting in better outcomes (Kneeland et al., 2016b). These findings, while preliminary, suggest that pressuring someone to exert control over their emotions may fail and unintentionally induce additional costs. Future research should investigate the conditions under which motivating others to regulate their emotions will achieve the best outcomes for them, as well as the conditions under which people are likely to err.

Additional Limitations and Future Directions

Study 2.1 showed that people who view their own emotions as controllable tended to judge others has having more control as well. This attribution of control predicted a decrease in sympathy and an increase in unsupportive behavior, replicating and extending past work showing a connection between dispositional theories of emotion (or happiness) control and reacting to others in a supportive or unsupportive manner (e.g., Tullett and Placks, 2016). This finding is striking in light of the observation that prior studies uniformly report good outcomes for people who judge their beliefs to be highly controllable (e.g., De Castella et al., 2013; Ford et al., 2017; Schroder et al., 2015; Tamir et al., 2007). This study is the first study to our knowledge to suggest that this trait comes with negative side effects – in this case, being less likely to be supportive of others. But while this finding is suggestive, it is not at all conclusive. It is possible that people who are better at regulating their emotions also tend to have other qualities that make them better at assessing others’ capacity for emotion regulation in real life, especially amongst their close others. Our study, which used hypothetical scenarios, would not have been able to account for this. Future work should investigate how one’s own capacity for emotion regulation affects one’s ability to successfully help (or motivate) others to regulate their own emotions.

We speculated that people are sensitive to what features of a person and situation make emotions more (or less) difficult to control. We found some support for this in the cue of emotion-situation calibration; however, a few questions remain about people’s reasoning about calibration and control. First, how do people reason about an emotion’s calibration? One possibility is that people base their calibration judgments based on associations about what feelings and situations often co-occur (e.g., Skerry & Saxe, 2015). Another possibility is that people reason about emotion calibration in virtue of how functional that emotion is in that moment – i.e., that stress is calibrated insofar as it is useful for motivating action (e.g., Troy et al., 2013). In our view, how people evaluate emotions as “rational” or “calibrated” is an open question and a valuable goal for future research. A second important question regards how accurate people’s judgments of others’ cognitive control over emotions are. People are notoriously poor at reasoning about others’ mental life. For instance, people often fail to appreciate how valuable or important certain things are to others (Pronin, Fleming, & Steffel, 2008) and people often think that others’ emotional experiences are less strong than their own (McFarland & Miller, 1990). And finally, work has recently shown that, because observers lack the direct experience that others have over their beliefs, that observers judge others as more able to voluntarily change beliefs than those individuals do (Chapter 3). Thus, even if a sufferer agrees her emotion is miscalibrated to her situation, she may disagree with an observer that she has voluntary control over it. Understanding how common these interpersonal conflicts are, and how people resolve them, remains a valuable goal for future research.

There are potentially many more cues people could use to derive attributions of emotion control. For instance, as time passes following a trauma, people are more capable of distracting themselves from, or finding meaning in, the event (Wilson & Gilbert, 2008; Wilson, Gilbert, & Centerbar, 2003). We should therefore predict that people will judge that someone mourning the loss of a loved one is more capable of regulating away that sadness several days or weeks after the death compared to immediately following it (e.g., as discussed by Whortman et al., 1988). Likewise, children and young adults improve the capacity to regulate their emotions as they grow (e.g., Band & Weisz, 1988; Fields & Prinz, 1997; Harris, Olthof, & Terwogt, 1981). Therefore, another prediction that falls out of the RER hypothesis is that parents will be less supportive of older-children's negative emotions compared to young children's negative emotions (calibrated to those children's emotional development). A valuable research goal is to further test how calibrated lay people are to features of individuals or situations that enable intrapersonal emotion regulation.

Conclusion

People seek out others to help regulate their negative emotional experiences. But when will their close friends and family acquiesce, and when will they react by making the sufferer feel even worse for feeling bad? We have showed that people reason about the degree of control that a sufferer has over his or her emotional reaction, and that they base their decision to show supportive or act unsupportively on this judgment. These control judgments predict people’s reactions towards suffering close others as well as people’s attitudes about costly policies designed to prevent emotional harm. We further showed that people arrive at these control judgments by reasoning about how well the emotion fits the situation, how rational the sufferer is, and how effective they themselves are at controlling their own emotions. These results are consistent with a form of emotion regulation *regulation* in which people expect and enforce others to regulate their own emotions if they can, and track features of the person or situation that enable emotion regulation success.

People judge others to have more control over beliefs than they themselves do

The language of belief is infused with attributions of control. People talk about what they and others *can* believe ("you can believe what you want, but if you ignore the rocks you'll be badly hurt"), what they *choose* to believe ("I choose to believe in the inherent intelligence and good sense of the average Malaysian voter," "One can choose to believe or not believe in God"), and what they *intend* or *decide* to believe ("He said he didn't know and I intend to believe him," "People are going to decide to believe what they want to believe").[7] The attributions that these locutions reflect play an important role in how people evaluate and react to others' beliefs. Indeed, in Chapter 1 I showed that people commonly attribute a high degree of intentional control to others over what they believe. That is, people incline towards judging that others (i) intentionally choose what they believe, (ii) have control over what they believe, and (iii) can choose to stop holding specific beliefs should they want to. Furthermore, just as with behavior, people appear to rely on these attributions of control when they evaluate belief holders. Individuals who attribute more intentional control to others over what they believe are more likely to blame those others for holding immoral or unjustified beliefs (Chapter 1). Thus, in keeping with the fundamental role that attributions of control play in determining how we explain and judge other people's behavior (Heider, 1958; Malle, Guglielmo, & Monroe,

[7] These examples were obtained from *The Corpus of News on the Web* (Davies, 2013).

2014; Skinner, 1996; Weiner, 1995), control attributions appear to occupy a similarly fundamental role when it comes to beliefs.

In the present studies, we investigate attributions of control over beliefs (also called “doxastic control”) for the self as compared with others. There are plausible theoretical reasons to predict that people would attribute more control to themselves over their beliefs than they attribute to others (over theirs); but there are also plausible reasons to predict the opposite pattern. Resolving this question therefore has important theoretical implications. On a more practical level, discrepant self-other judgments about doxastic control have the potential to exacerbate real-world disagreements over discordant beliefs. For instance, people often feel personally affronted when others do not share their beliefs (e.g., Golman, Loewenstein, Moene, & Zarri, 2016). If they also judge themselves to have less (or more) control over changing their beliefs than others judge them to have, then this may lead to discrepant expectations about which party can choose to change their minds, thus potentially amplifying the original conflict. Both factors point to the relevance of understanding whether, and why, people judge that they and others have different levels of control over what they judge to be true.

As noted above, two divergent predictions emerge from past research. One line of research predicts that people should tend to judge themselves to have more control over their own beliefs than others have over theirs because, to most people, control is desirable, and people often self-enhance desirable properties. People have a strong preference to feel and exert control and react negatively to feeling a loss of it (e.g., Brehm, 1966; Burger & Cooper, 1979; Kelley, 1971; Seligman, 1974, 1975; Wortman & Brehm, 1975). Indeed, the desire for control has been described as one of the strongest human motivations (Gebhardt & Brosschot, 2002; see also Bandura, 1977; Deci & Ryan, 1986; White, 1959). As a consequence, people tend to over-attribute control to themselves, inflating how much control they think they have over many things in their life. Indeed, past work suggests that attributions of control are readily biased by motivational concerns (e.g., Alicke, 2000; Burger, 1986; Clark et al., 2014; Mazzocco, Alicke, & Davis, 2004; Miller & Norman, 1975). For instance, the so-called "illusion of control" – whereby people attribute control to themselves over things they in fact have no control over (Langer, 1975; but see Gino, Sharek, & Moore, 2011) – appears especially pronounced in individuals who have a strong desire for control (Burger, 1986).

Because the desire for control pertains to the self and not to others, we would therefore expect people to inflate self-directed, but not other-directed, attributions. Consistent with this reasoning, several studies have found that people attribute to themselves greater control over their own actions than they grant to others (Pronin & Kugler, 2010). In particular, people regard their behavior as driven more by their own intentions and desires than the same behaviors performed by their roommates (Pronin & Kugler, 2010, Study 4; see also Miller & Norman, 1975). Therefore, if people reason about their beliefs in the same way that they reason about their behavior, they should grant themselves more volitional control over their beliefs than they grant to others (over theirs).

Yet there is also reason to postulate precisely the opposite prediction, namely, that people will attribute *less* belief control to themselves than they attribute to others. Our argument proceeds as follows: Unlike actions, beliefs are typically experienced as uncontrollable on account of internal, psychological constraints on belief change.

However, because people tend to have difficulty reasoning about, and fully accounting for, the hidden, psychological constraints operating on others, they should routinely fail to account for these constraints when attributing control to others. Instead, they rely on a default, unreflective judgment that beliefs, like behaviors, are generally controllable, which results in their attributing high degrees of control over beliefs to other people. The combination of these factors should yield a self-other discrepancy, such that believers attribute to themselves less control over their beliefs than they attribute to others. We motivate this line of reasoning below, with two key premises.

Our first premise is that beliefs – more so than actions – are subject to psychological constraints that limit people's ability to change them voluntarily (James, 1937). One major source of constraint is the objective evidence that impinges upon people's beliefs. Indeed, existing work that directly investigates belief formation and change suggests that beliefs are partially outside people's voluntary control, precisely because of these evidentiary constraints. In essence, while people can indirectly influence the quality of their beliefs, including how rational and justified those beliefs are (e.g., by exposing themselves to new information, or by deliberating in specific ways; Baron, 2008; Haran, Ritov, & Mellers, 2013; Stanovich & West, 1997; Webster & Kruglanski, 1994), they cannot simply adopt whatever belief they want to (Epley & Gilovich, 2016; Sloman, Fernbach, & Hagmeyer, 2010). For instance, when presented with strong arguments in favor of a proposition, people tend to change their beliefs, even when they would prefer not to (Petty & Cacioppo, 1986; Wood & Porter, 2016).[8]

[8] This conclusion is not undermined by the phenomenon of motivated reasoning, as it might seem to be at first. While there is widespread agreement that people sometimes

Anecdotal evidence suggests that people sometimes do indeed experience their beliefs as constrained, which leads them to view themselves as having low control over their beliefs. Consider this passage in William James's essay, *The Will to Believe*, in which he reflects on this evidentiary constraint on belief:

> Can we, by any effort of our will, or by any strength of wish that it were true, believe ourselves well and about when we are roaring with rheumatism in bed, or feel certain that the sum of the two one-dollar bills in our pocket must be a hundred dollars? We can say any of these things, but we are absolutely impotent to believe them (p. 5, 1937).

In this passage, James claims that, because he has strong evidence in favor of certain beliefs (e.g., that he has two dollars in his pocket), he is literally unable to form a contrary belief. Similar anecdotes have been provided by other scholars who introspect on their own beliefs (e.g., Alston, 1988; Epley & Gilovich, 2016; Pascal, 1852, see Turri et al., 2017 for a review). Thus, it appears that confrontation with evidence limits the magnitude and scope of voluntary belief change, and does so in a way that gives rise to a *feeling* of belief constraint discoverable through introspection (Alston, 1987; James, 1937; Kunda, 1990).

reason in motivated (i.e., biased) ways (Baumeister & Newman, 1994; Kunda, 1990), the existence of this phenomenon does not imply that people have conscious volitional control over their beliefs. In fact, motivated reasoning is likely to work best when it bypasses the will, with the relevant motivations affecting the kinds of information that people consider, rather than operating directly via the will to control final beliefs states (Baumeister & Newman, 1994; Epley & Gilovich, 2016). We discuss the relevance of motivated reasoning, as well as the claim that beliefs in fact are uncontrollable, in the General Discussion.

People may feel evidentially constrained even when there is little objective evidence available to them. A core postulate of the well-known theory of naïve realism is that people assume that they "see entities and events as they are in objective reality" and that their "social attitudes, beliefs, preferences, priorities, and the like follow from a relatively dispassionate, unbiased, and essentially 'unmediated' apprehension of the information or evidence at hand" (Ross & Ward, 1996, p. 110; see also Griffin & Ross, 1991). This tendency – separate from the actual state of the evidence with regard to any particular belief – could exacerbate the feeling of belief constraint.

Finally, the feeling of constraint may be further amplified by non-evidentiary factors. Many beliefs are formed through unconscious or a-rational processes that people may have little insight to (Nisbet & Wilson, 1977). For instance, repeated exposure to some stimulus may influence downstream beliefs that something is preferable, safe, or of high quality (Zajonc, 1980). Ordinary people may be genuinely unable to override these attitudes, or may simply not understand how to by virtue of not having access to how they came about in the first place (Wilson & Brekke, 1994). In sum, a mixture of both evidentiary and non-evidentiary factors could jointly contribute to the feeling people have that their beliefs are constrained. To date, however, no work has examined whether lay people actually experience their own beliefs as outside of their control, which was a major purpose of our investigations.

Our second premise is that people fail to appreciate that others suffer this same sense of constraint over their beliefs. This idea derives from a broader difficulty people have in appreciating others' inner experiences (Pronin, 2009), stemming from the fact that people do not directly experience others' mental states but have to infer them indirectly (Jones & Nisbett, 1972). As illustrative of this difficulty, people judge that others have less complex mental experiences than they themselves do, fail to appreciate the subjective importance of others' experiences, and judge others' emotions as less intense than their own (Johnson, 1987; McFarland & Miller, 1990; Miller & McFarland, 1987; Pronin, Kruger, Savtisky, & Ross, 2001; Pronin, Fleming, & Steffel, 2008). Furthermore, when people do not directly experience some emotion or psychological pressure, they often fail to account for it when predicting and explaining behavior. Failures to appreciate others' internal constraints result in biased attributions, as well as errors in predicting others' behavior (e.g., Bierbrauer, 1979; Jones & Harris, 1968; see Gilbert & Malone, 1995, for a review). Moreover, people often fail to account for psychological constraints operating on *themselves* if they are not directly experiencing them in the moment, leading to similar prediction errors (e.g., Gilbert, Gill, & Wilson, 2002; Loewenstein, 1996; Van Boven & Loewenstein, 2003). Based on this background research, we predicted that people will insufficiently incorporate others' felt experience of low belief control, attributing more control than those others attribute to themselves.

To summarize, if people's ability to alter their beliefs is genuinely constrained – as it appears to be – and if it is constrained by forces that are not directly observable in others', such as the evidence perceived in favor of a particular proposition – as it also appears to be – then we should expect observers to have difficulty incorporating these internal constraints when judging others' control over their beliefs. Actors, however, should readily encounter those constraints when they introspect on their own beliefs, and should therefore attribute lower control to themselves over their beliefs. As a consequence, we should expect people routinely to judge themselves as having less control over their own beliefs than others have over theirs.

Overview of Studies.

The present set of studies sought to test these predictions. No work that we are aware of has measured whether lay people judge that they have control over their own beliefs; similarly, no work has compared self and other-directed ratings of belief control. Our studies address these questions, thereby enabling a test of the two competing theories described above. Based on the reasoning outlined above, our prediction was that, when considering specific, concrete beliefs, people would attribute to themselves less voluntary control over their beliefs than they would attribute to others. However, we remained open to the possibility that the alternative prediction would instead prove correct (more belief control attributed to self than other), and the studies were capable of revealing this.

We conducted five studies to address these issues. In Study 3.1, we find that for opposing beliefs on important social issues, people reliably judge themselves to have less ability to change their beliefs than others have. In Study 3.2, we find that this effect generalizes to a case in which self and other hold the same belief (rather than opposing beliefs). In Study 3.3, we tested an alternative account of our findings, namely that the self-other discrepancy for beliefs arises because subjects think that it would be bad (or look bad) to say that they can change their beliefs (as might be predicted by some theories of self-enhancement or self-presentation).

In Studies 3.4 and 3.5 we examined whether the self-other discrepancy is attenuated when people reason abstractly about their own and others' doxastic control. In

Study 3.4, we find that people attribute to themselves more control when considering their belief control in the abstract, than when considering specific beliefs that they hold. And in Study 3.5, we find that the self-other difference in control occurs only when people consider specific beliefs; it is fully attenuated when people consider their own and others' control over beliefs in general. These latter findings are directly predicted by the theoretical reasoning outlined previously, which asserts that it is the introspective experience of low control that drives down self-directed attributions of control relative to other-directed attributions.

Study 3.1

Study 3.1 investigated whether people judge their own ability to change a belief about an important social issue differently from another person's ability to change their opposing belief on the same issue. We examined subjects' beliefs on four topics: (a) whether God exists (God), (b) whether genetically modified foods should be prohibited (GMF), (c) whether government regulation is the best way to address global climate change (Climate), and (d) whether social media has had a negative overall impact on dating (Social Media). We selected these topics because they reflect timely and important social issues over which people frequently disagree (Pew Research Center 2015, 2016a, 2016b, 2016c).

Our primary prediction was that judgments of belief control for the self would be lower than corresponding control judgments for the other person. The study also contained an exploratory component. We were interested in whether this predicted discrepancy would apply across two distinct judgments of belief control. The two judgments concerned (1) whether the agent (self or other) *could choose to change the belief if they wanted to*, and (2) whether the agent (self or other) *intentionally chose* to hold a particular belief. Judging whether one could choose to change a belief involves considering one's current ability to intentionally bring about belief change[9]. It captures the notion of control in past scholars' introspective accounts of low belief control (see the William James example provided earlier), and it is also consistent with the notion of control in metacognition recognition research (e.g., Nelson & Narens, 1990). By contrast, judging that one intentionally chose a belief is a retrospective judgment that involves recalling one's history of believing some idea, including whether one had a desire to adopt the belief in the first place (Malle & Knobe, 1997). We included it because we considered it possible that such retrospective judgments of intentional choice would similarly yield a self-other discrepancy, with ratings of the self's intentionality lower than those of others' intentionality.

Method

Participants. We recruited 394 people (mean age = 37, 184 reported Female) from Amazon's Mechanical Turk system to participate in the experiment.

Design. We used a 2x2x4 design, investigating attribution target (self vs. other), control measure (chose vs. change), and belief (God vs. GMF vs. Climate vs. Social Media). Subjects were randomly assigned to make judgments either about their own or

[9] This measure comprised our primary dependent variable in the full set of studies we report, although at the time we ran Study 3.1 and we were not sure how it would differ from the choice measure.

others' belief control (between subjects), but they responded to all four belief contents and both control measures (within subjects).

Procedure. At the beginning of the study, subjects reported their current belief on four topics, God, GMF, Climate, and Social Media. They did so by choosing which of two opposing statements they agreed with on each issue (e.g., God exists vs. God does not exist, genetic modification should be prohibited vs. genetic modification should not be prohibited; see Appendix J for full text of stimuli). Subjects were then randomly assigned to respond to follow-up questions about either their own control over these beliefs (self), or alternatively, another person's control (other) over beliefs opposite to those held by the self. Subjects indicated whether they (or the other person) deliberately chose to hold each of the four beliefs (chose), as well as whether they (or the other) could choose to believe the opposite belief (change).

Which statements subjects were presented with varied depending on what subjects indicated they believed at the beginning of the study. In the self condition, if a subject initially indicated that he or she believed that genetically modified foods should be prohibited, they would then have rated their agreement with the following chose statement, "I deliberately chose to believe that genetically modified foods should be prohibited," and with the following change statement, "If I wanted to, I could choose to believe that genetically modified foods should not be prohibited." However, if the subject originally indicated that they believed that genetically modified foods should not be prohibited, then they would instead have rated their agreement with, "I deliberately chose to believe that genetically modified foods should not be prohibited," and "If I wanted to, I could choose to believe that genetically modified foods should be prohibited."

In the other condition, before rating control, participants were given the following instructions (paragraph breaks indicated by "//"):

> We are now going to ask you a series of questions about other people who, in a prior study we conducted, indicated what they believed about each of these four topics. // We are keeping it confidential who they were, just as all data we collect is kept confidential, so try to imagine another Mechanical Turk worker, similar to yourself, who holds the belief we describe. // In each case, you will be reading about a person similar to you but who holds an attitude that you do not hold.

For each of the four beliefs, participants then saw a statement like the following (bold in original text):

> A participant from a previous experiment indicated that he/she "**believes that genetically modified foods should <u>not</u> be prohibited**". You indicated that you believe the opposite. // Please indicate your agreement with the following statements. (Even if you are not certain of the answer, please indicate what you think is most likely.)

The specific content was matched to be the opposite the subject's own belief. The chose and change questions used wording similar to that used in the self condition (see above), modified as needed for the other condition. For instance, in the case above, subjects rated their agreement with, "This person deliberately chose to believe that genetically modified food should <u>not</u> be prohibited," and "If this person wanted to, he/she could choose to believe that genetically modified foods should be prohibited."

All ratings were made on 1-7 rating scales with 1 labeled "completely disagree" and 7 labeled "completely agree." Subjects in both the self and other conditions responded to control questions for all four beliefs, which were shown on separate screens in an order randomly set for each participant. At the end of the study subjects indicated their sex, age, and were debriefed. No other measures were collected.

Results

As planned, we ran a linear mixed-effect model regressing agreement ratings on attribution target (self vs. other), control measure (chose vs. change), and their interaction. The model also included a random by-subject and by-belief content intercepts as well as a random by-subject and by-belief content slopes for the effect of control measure[10]. We observed significant effects of Target, ($b = -0.53$, $SE = 0.12$, $t = -4.33$, $p < 0.001$, $R_{(m)}^2 = 0.02$), and control measure ($b = 0.81$, $SE = 0.18$, $t = 4.60$, $p < 0.001$, $R_{(m)}^2 = 0.04$), as well as their interaction ($b = 1.42$, $SE = 0.17$, $t = 8.20$, $p < 0.001$, $R_{(m)}^2 = 0.04$). See Figure 3.1.

[10] In Studies 3.1 and 3.2 we computed linear mixed-effect models using the lme4 (Bates, Mächler, Bolker, & Walker, 2015) package in the R computing environment. For effect sizes, we calculated partial-R^2 ($R_{(m)}^2$) for each fixed effect using the r2glmm package (Jaeger, 2017) which implements the approach suggested by Nakagawa and Schielzeth (2013).

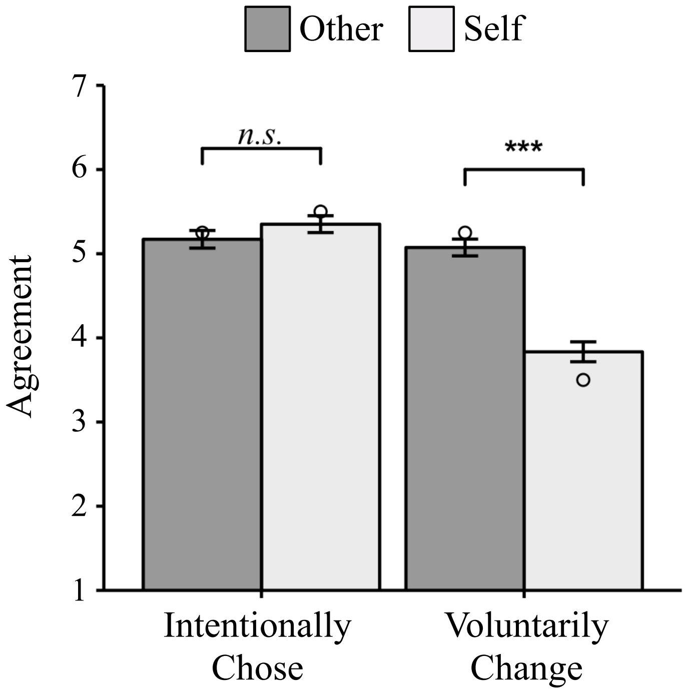

Figure 3.1. Means and standard errors from participant responses in Study 3.1. Circles represent median value.

Tests of simple effects revealed that, in accordance with our main prediction, subjects' judgments of their own ability to voluntarily change their beliefs (*M* = 3.83, *SD* = 2.12) were significantly lower than corresponding judgments of others' ability to do so (*M* = 5.08, *SD* = 1.73), *b* = -1.24, *SE* = 0.15, *t* = -8.00, *p* < 0.001, $R_{(m)}^2$ = 0.05. However, subjects did not report *choosing* their beliefs any more or less than others (*b* = 0.18, *SE* = 0.14, *t* = 1.23, *p* = 0.218, $R_{(m)}^2 < 0.01$).[11] When analyzed separately, all four beliefs

[11] In order to be more precise regarding this finding of a lack of difference, we averaged *choose* values for each subject and conducted an equivalence test using the two one-sided t-test procedure (provided by the TOSTER package, Lakens, 2017). The equivalence test was significant, *t*(391.84) = 1.98, *p* = 0.0242, given equivalence bounds of *d* = -0.30 and *d* = .30 and an alpha of 0.05. We should conclude from this that, if there truly is a difference in self and other attributions of intentional choosing beliefs, it must be smaller

revealed the same pattern of results (see Table 3.1 for means and standard deviations; see Figure 3.2 for change ratings).

Table 3.1
Mean (and standard deviation) of judgments about each belief in Studies 3.1 and 3.2.

Study	Control Measure	Attribution Target	Climate	GMF	God	Social Media
Study 3.1	Chose	Other	5.11 (1.81)	5.35 (1.72)	5.21 (1.93)	5.03 (1.69)
		Self	5.36 (1.68)	5.41 (1.70)	5.41 (2.01)	5.23 (1.77)
	Change	Other	5.15 (1.64)	5.22 (1.67)	4.93 (1.95)	5.01 (1.63)
		Self	3.76 (2.03)	4.09 (2.02)	3.21 (2.31)	4.29 (1.94)
Study 3.2	Chose	Other	5.06 (1.86)	4.99 (1.95)	5.38 (1.83)	4.62 (1.97)
		Self	5.37 (1.72)	4.93 (2.03)	5.49 (2.14)	4.90 (1.90)
	Change	Other	4.48 (2.04)	4.66 (2.08)	4.29 (2.26)	4.93 (1.82)
		Self	3.97 (2.21)	4.32 (2.14)	3.57 (2.39)	4.53 (1.97)

Note. Ratings made on a 1-7 scale (1 = ‘completely disagree’; 7 = ‘completely agree’).

We also conducted exploratory analyses investigating whether the self-other difference in judgments of voluntary change held when separately analyzing those who endorsed the statements (i.e., those who *did* believe in God, or that GMFs should be prohibited, etc.), and those who held the opposite, “negative” viewpoint (e.g., those who believed that God *does not* exist, or that GMFs should *not* be prohibited, etc.; see Figure 3.3). For both groups, subjects’ judgments of their own ability to voluntarily change their beliefs (endorse: *M* = 4.64, *SD* = 2.13; non-endorse: *M* = 4.55, *SD* = 2.08) were lower than judgments of others’ ability to change theirs (endorse: *M* = 5.21, *SD* = 1.73; non-endorse: *M* = 5.02, *SD* = 1.79). A linear mixed-effect model regressing change ratings on the attribution target, with a random intercept for subject, confirmed the presence of a

d = .3 or we would have found it. We use this procedure for all subsequent equivalence tests reported in the paper.

self-other difference for "endorsers" (b = -1.55, SE = 0.17, t = -9.15, p < 0.001, $R_{(m)}^2$ = 0.14) and for "non-endorsers" (b = -0.91, SE = 0.18, t = -4.97, p < 0.001, $R_{(m)}^2$ = 0.05).

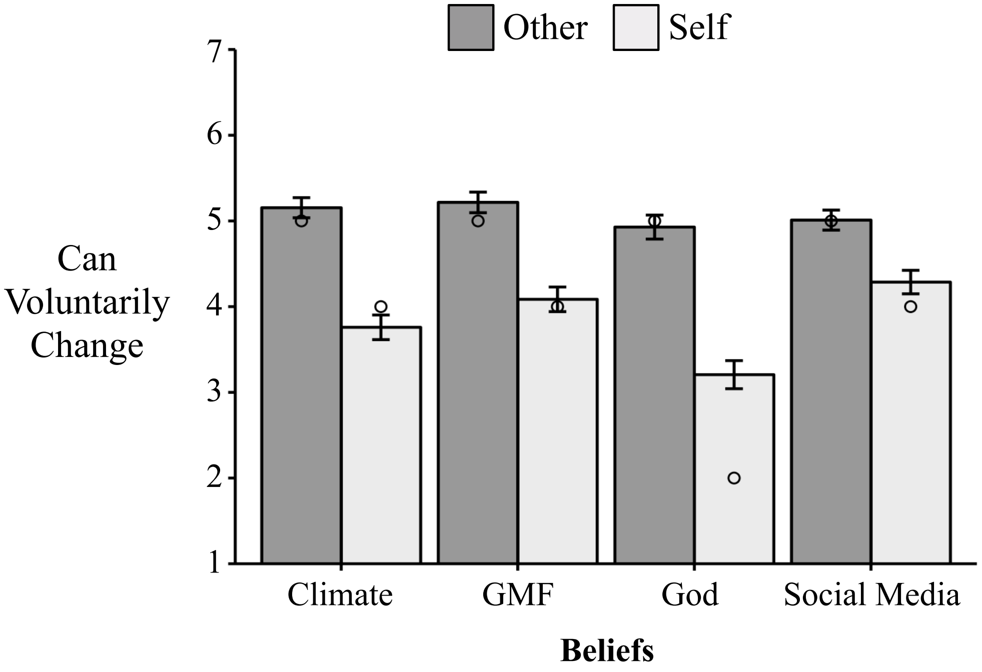

Figure 3.2. Mean (and standard error) for subject's agreement ratings in response to change question, across self and other conditions, for each of the four beliefs. Circles represent median values.

Discussion

Study 3.1 showed that people judged themselves less capable than others of changing their beliefs. This finding replicated across all four belief statements investigated in this study, and did so regardless of whether subjects endorsed the statements or not. This finding points to the possibility that there are *two* sources of disagreement in cases of everyday belief conflict: individuals with opposing beliefs disagree not only about the matter at hand, but also about who between them could choose to change their mind. We address the implications of this finding in the General Discussion.

In contrast with change ratings, we observed no difference between judgments of whether the self or the other person intentionally chose to have the beliefs in question. This null effect was not directly predicted, nor was the difference between change and choose ratings. However, as we noted in the Introduction to this study, the "change" measure of control is most consistent with operationalizations of control in the meta-cognition literature, as well as with prior anecdotal reports of low belief control (e.g., William James's account). One post-hoc explanation for the observed difference is that, because it is focused in the present (or the immediate future), only the change measure directly confronts people with the limits they face when trying to control a given belief, whereas the intentional choice measure, being retrospective (did you choose this belief in the past?), evokes these limits much more indirectly (as well as being subject to memory loss concerning any given belief formation process). Given that the theory predicting lowered self-ratings of control hinges on whether subjects directly experience the limits to their own control, this difference in the focus of the two questions may account for the difference in the pattern of ratings. Consistent with this idea, control ratings were lower overall for the change measure than for the choice measure, as a function of the lower ratings in the self-change condition, specifically. Regardless of the explanation, in Study 3.2 we examined whether this difference between the two measures replicated.

The self-other discrepancy for change judgments provides initial support for the theory that people view their own beliefs as less controllable than others', while also highlighting a potentially important dynamic between people who hold opposing attitudes. However, the fact that subjects only judged someone who held an opposing belief leaves open the possibility that the discrepancy is limited to cases of disagreement, rather than reflecting a more general self-other difference. In particular, people might judge that the disagreeing other has an incorrect belief, and that incorrect beliefs are more changeable than correct ones, not that other people generally have more control over their beliefs. If the self-other discrepancy is truly general, it would need to replicate in cases where the other person holds the same belief as the self. We therefore tested this in Study 3.2.

Study 3.2

Method

Participants. 198 people (mean age = 39; 112 reported Female) recruited from Amazon's Mechanical Turk platform participated in the experiment.

Design and Procedures. We replicated the design of Study 3.1 for Study 3.2, crossing attribution target (self vs. other), control measure (chose vs. change), and belief (God, GMF, Climate, Social Media). After first indicating their own beliefs, subjects were randomly assigned to respond either to questions about their own, or another person's beliefs, and they answered both control questions for all four beliefs.

Subjects first rated their agreement with four statements that were adapted from the topics used in Study 3.1 (e.g., "God exists"; see Appendix K for full text of all items). The four statements were presented in a new random order for each subject, and each rating was made on a 6-point rating scale with the following options, in order: "strongly disagree," "disagree," "somewhat disagree," "somewhat agree," "agree," and "strongly agree." Subjects were next randomly assigned to either the self or the other condition, and answered follow-up control questions about each of the four beliefs.

In the self condition, subjects were first reminded of what they had just reported believing, and were then asked to indicate their agreement with claims that they "deliberately chose" and "could choose to/to not believe" the earlier statements they had endorsed. For instance, those who indicated that they "strongly agreed" with the statement, "genetically modified foods should be prohibited," were reminded, "You indicated that you **strongly agree** with the statement "**Genetically modified foods should be prohibited."** // Please indicate your agreement with the statements below." (bold original). The two statements pertained to deliberate choice (e.g., "I **deliberately chose to believe** that genetically modified foods should be prohibited") and the ability to choose *not* to believe the statement (e.g., "If I wanted to, **I could choose <u>not</u> to believe** that genetically modified foods should be prohibited"), in that order. For cases in which subjects had first indicated disagreement, the subsequent statement about choosing to believe was modified to "deliberately chose not to believe," and the voluntary change statement referred to believing the proposition.

In the other condition, subjects were provided instructions indicating that they would answer questions about a person who believed the same thing that they did:

> We are now going to ask you a series of questions about other people who hold similar beliefs to you - specifically people who responded the same way to these questions in earlier studies. // In each case, we are reporting someone's belief that we measured in a prior study we conducted (though of course we are keeping it confidential who they were, just as all data we collect is kept confidential). // For

> each of the following questions, try to imagine another Mechanical Turk worker, similar to yourself, who holds the belief we describe.

To illustrate, subjects who indicated that they "strongly agree" with the GMF statement were presented with the following prompt, "Another mechanical turk worker, from a prior study we conducted, indicated that they **strongly agree** with the statement "**Genetically modified foods should be prohibited**." These subjects then indicated their agreement with the chose and change questions: "This person **deliberately chose to believe** that genetically modified foods should be prohibited," and "If this person wanted to, **he/she could choose <u>not</u> to believe** that genetically modified foods should be prohibited," respectively. As in the self condition, the prompts were modified to match subjects' initial agreement or disagreement with each statement.

Subjects rated their agreement with the chose and change questions on 7-point scales with 1 indicating "completely disagree" and 7 indicating "completely agree." The four items (each consisting of a pair of questions) were presented on separate pages and in a random order for each subject. At the end of the study, participants reported their sex and age before being debriefed. No other measures were collected.

Results

As planned, we followed the same analysis procedure from Study 3.1, regressing agreement ratings on control type, attribution target, and the interaction of control type and attribution target, using a linear mixed-effects model with random intercepts for subject and belief content, and random by-subject and by-belief slopes for the effect of control type. These analyses revealed a significant effect of control type, such that chose ratings were higher ($M = 5.09$, $SD = 1.51$) than change ratings ($M = 4.34$, $SD = 1.73$), $b = 0.75$, $SE = 0.14$, $t = 5.52$, $p < 0.001$, $R_{(m)}^2 = 0.03$). There was no main effect of target (Self: $M = 4.80$, $SD = 1.64$; Other: $M = 4.64$, $SD = 1.69$), $b = -0.17$, $SE = 0.19$, $t = -0.90$, $p = 0.37$, $R_{(m)}^2 < 0.01$, but there was a significant interaction between control type and target, just as there had been in Study 3.1, $b = 0.66$, $SE = 0.27$, $t = 2.41$, $p = 0.016$, $R_{(m)}^2 = 0.01$.

A test of simple effects confirmed our prediction that ratings of change were lower for self ($M = 4.10$, $SD = 1.72$) than for other ($M = 4.59$, $SD = 1.72$), $b = -0.49$, $SE = 0.24$, $t = -2.02$, $p = 0.043$, $R_{(m)}^2 = 0.01$, with no corresponding difference between self ($M = 5.17$, $SD = 1.48$) and other ($M = 5.01$, $SD = 1.54$) for chose ratings, $b = 0.16$, $SE = 0.22$, $t = 0.75$, $p = 0.453$, $R_{(m)}^2 < 0.01$ (see Figure 3.3). A follow-up equivalence test with bounds $d = -0.40$ and $d = 0.40$ was significant, $t(195.81) = 2.23$, $p = 0.013$, suggesting that if there was a self other discrepancy for intentioncal choice ratings, it must be smaller than $d = 0.40$. Change ratings were lower than choose ratings for both self, $b = 1.08$, $SE = 0.19$, $t = 5.61$, $p < 0.001$, $R_{(m)}^2 = 0.03$, and other, $b = 0.42$, $SE = 0.19$, $t = 2.19$, $p = 0.028$, $R_{(m)}^2 = 0.01$[12].

[12] We also repeated the analysis from Study 3.1 investigating differences between self and other for each belief individually. The results were less consistent than we observed in Study 3.1. Self-directed ratings were lower than other ratings for God, $t(195.42) = -2.20$, $p = 0.029$. However, they were not significantly different for Media, $t(194.77) = -1.71$, $p = 0.089$, GMF, $t(195.81) = -1.11$, $p = 0.268$, or Climate, $t(194.76) = -1.50$, $p = 0.136$.

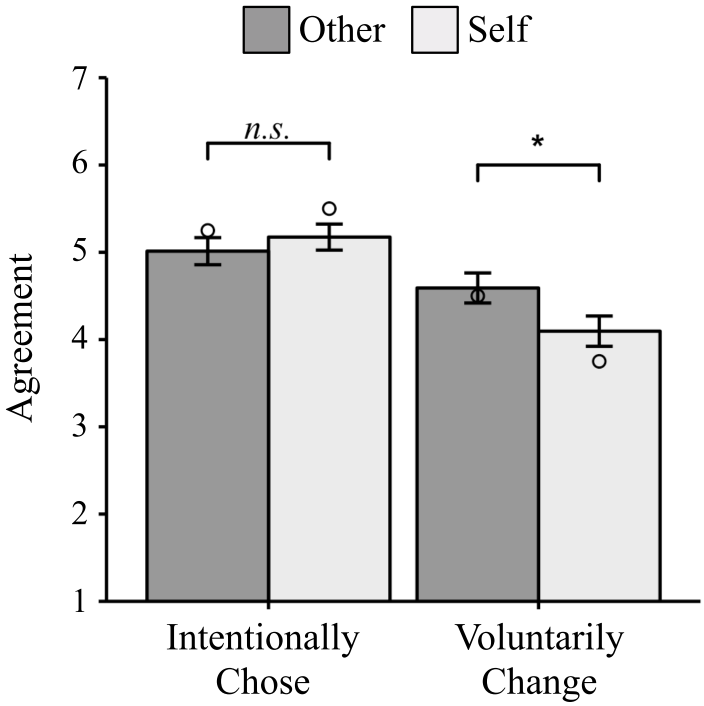

Figure 3.3. Mean ratings (and standard errors) for Chose and Change judgments Self vs. Other in Study 3.2 (matched beliefs). Circles represent median values.

Discussion

Study 3.2 replicated Study 3.1's findings that people judge others to have a greater ability to voluntarily change their beliefs than they themselves do. In Study 3.2, this result occurred even though subjects judged another person's ability to stop believing a mutually shared belief rather than an opposing belief. This finding suggests that there may be a general self-other discrepancy in attributions of control over beliefs, rather than the difference being limited only to cases of disagreement. As in Study 3.1, this difference occurred only for judgments of the voluntary ability to change one's beliefs and did not occur for judgments of intentional choice; thus we have further evidence that the self-other difference is specific to judgments of voluntary change. For this reason, we focused only on judgments of voluntary change in the subsequent studies, and return briefly to this issue in the General Discussion.

The findings thus far are consistent with the idea that people's unique introspective access to the constraints on their own beliefs causes them to rate their own belief control lower than that of others. But there are some alternative explanations for this discrepancy that need to be addressed. One in particular is that people may regard voluntarily changing their beliefs (especially without exposure to new, justifying information) as wrong or counter-normative, which in turn affects ratings of their own belief control. This idea is encapsulated by William James, who writes, "the talk of believing by our volition… is worse than silly, it is vile" (p. 7, 1937; see also discussion in Clifford, 1877). Corroborating this perspective, recent research has indeed shown that some people regard adhering to the norms of rationality as a moral issue (Ståhl, Zaal, & Skitka, 2016). And, since people tend to regard the majority of their beliefs as reasonable and justified (Pronin, Gilovich & Ross, 2004; Ross & Ward, 1996), voluntarily changing these beliefs may therefore be seen by many subjects as unjustified or irresponsible, and therefore, as morally questionable. Accordingly, this might explain why people are reluctant to grant themselves the capacity to exert voluntary control over their beliefs.

The putative badness of voluntary belief change raises two distinct alternative mechanisms for the findings so far. First, people may privately judge that they are less able to voluntarily change their beliefs on the grounds that, as generally good people, they are less capable than others of immoral behavior. Supporting this idea, prior findings suggest that people generally hold a more favorable moral view of themselves compared to others (e.g., Alicke, 1985; Allison, Messick & Goethals, 1989). Second, subjects' judgments may reflect their desire to present themselves in a good light to the experimenter. On this account, subjects judge it as reputation enhancing to say that they could not perform some unvirtuous behavior – a motivation that would depress reports of their own control over their beliefs, but not others' control. If either of these alternatives accounts for the asymmetry reported above, then the results would reflect existing and well-established biases.

Accordingly, Study 3.3 tested whether the tendency to regard one's own beliefs as less controllable than others' beliefs is explained by a mechanism specific to beliefs – as our theorizing postulates – or whether it might instead reflect more general self-presentational or self-enhancement concerns. We compared people's judgments of their ability to voluntarily change a belief with their judgments of their ability to voluntarily perform an immoral behavior. Our hypothesis is that when people contemplate voluntarily changing a specific belief, the apparent psychological constraints on belief change loom large in their thinking (see Introduction). These constraints are less vividly appreciated for others, however, causing a self-other discrepancy in ratings of belief control. Reasoning about a hypothetical behavior, by contrast, should not evoke the same feelings of constraint. Indeed, prior work attests to people's frequent failure to simulate various visceral constraints on their behaviors, leading them to underappreciate the uncontrollable nature of those behaviors (Loewenstein, 1996). Because this neglect should be equally lacking for both self and other (Epley & Waytz, 2010), our theorizing predicts an attenuated self-other difference for behavior relative to belief. In contrast, if self-ratings of belief control are depressed because people regard belief change as bad, then subjects should similarly depress reports that they could choose to perform an (equivalently bad or worse) immoral behavior.

Study 3.3

In Study 3.3, subjects reported how much control either they or a close other would have either to believe that a prototypically immoral act was not immoral, or to perform that very same immoral act. In post-tests, performing the immoral act was rated as worse than the holding the immoral belief (see below). The study therefore represents a conservative test, since the alternative explanations under consideration hinge on the idea that people are unwilling to report voluntary control over beliefs because it is wrong or socially undesirable. If the earlier results instead reflect a belief-specific asymmetry (at least in part), then the self-other discrepancy for belief control should be larger than that observed for behavior control.

Study 3.3 departed from Studies 3.1 and 3.2 in another important way. Whereas in Studies 3.1 and 3.2, subjects rated a distant other's belief control, in Study 3.3 we prompted subjects to make judgments about someone very close to them. Past research has shown that people more readily project their own mental states to close rather than distant others, and that they are also more inclined to adopt the perspective of liked versus disliked others (Epley & Waytz, 2010). Thus, Study 3.3 represents a conservative test in this way as well, since a self-other discrepancy should be less likely to occur for close others.

Method

Participants. In order to make meaningful self-other comparisons (and also meaningful comparisons between beliefs and behavior), subjects had to report both that they thought the action in question was wrong, and that their nominated close other also believed this. Subjects who did not do so were excluded, and we preregistered this exclusion plan.

Data collection occurred in two phases. In the first phase, we recruited 549 people from Amazon's Mechanical Turk. 55 of these subjects (10%) did not qualify for the study because they failed one of the two exclusion criteria above. To reach our recruitment target of 500, we recruited another nine people (also from Amazon's Mechanical Turk). Of these, 1 failed our criterion for inclusion, yielding a final sample of 502 subjects (mean age = 36.2, 258 reported Female).

Design. The study had a 2x2 mixed between-within design, with the target of control attributions (self vs. other) manipulated between-subjects and the type of behavior (belief vs. act) manipulated within-subjects. Subjects were randomly assigned to either the self or other condition, and in each case, responded to control judgments about changing a moral belief and performing a corresponding immoral behavior (in random order).

Procedure. At the beginning of the study, subjects read a short description of a prototypical immoral behavior:

Sometimes people take advantage of another person's costly mistake. Specifically, sometimes a person will see that another person walking ahead of them has

> *dropped $20 on the ground but, instead of returning the money, the person behind will just keep it.*

After reading about this behavior, subjects reported whether they "Agree" or "Disagree" with the statement, "I believe that in this situation it is wrong to keep the $20 instead of returning it to its original owner." Next, participants were asked to think of someone close to them, such as a best friend, romantic partner, or spouse, and to type out that person's initials. Once they had done so, they were asked whether they agree that, "The person whose initials I typed above believes that in this situation it is wrong to keep the $20 instead of returning it to its owner." Subjects could select either, "Agree, this person believes that in this situation it is wrong to keep the $20 instead of returning it to its owner," or "Disagree, this person believes that in this situation it is not wrong to keep the $20 instead of returning it to its owner."

If subjects indicated either that they believed that keeping the $20 was not wrong, or that the close other they nominated did not believe that it is wrong, then they were not selected to continue in the study. Instead, they were redirected to the short demographics questionnaire (described below), then debriefed, and paid in full for participating. We excluded these subjects so that all subjects were making control judgments about a belief they shared with their close other. This ensured that we did not re-introduce a possible confound we eliminated in Study 3.2.

Subjects who indicated that both they and their close other believed the target act was wrong were randomly assigned to answer follow-up questions about either themselves or their close other. In the self condition, subjects indicated their agreement with a statement about whether they could choose to perform the immoral act, "In this situation, if I wanted to, I could choose to keep the $20 instead of returning it to its original owner," as well as a statement about their ability to believe otherwise about the moral status of this act, "If I wanted to, I could choose to believe that keeping the $20 instead of returning it to its owner is not wrong." The order of these statements was counter-balanced, and agreement was assessed using a 7-point rating scale (1 = "Completely disagree; 7 = "Completely agree"). In the other conditions, the statements were altered by dynamically inserting the initials of the person the subject had nominated. For instance, if I nominated my advisor, Geoff Goodwin, the statements would have read, "In this situation, if GG wanted to, GG could choose to keep the $20 instead of returning it to its original owner," and "If GG wanted to, GG could choose to believe that keeping the $20 instead of returning it to its owner is not wrong," for the act and belief conditions, respectively. For each statement, subjects were instructed to provide their answer without considering how much they (or the close other) actually would want to do or believe such a thing. Subjects answered each question on a different screen.

At the end of the study, subjects indicated their age and sex, and then were debriefed. No other measures were collected.

Results

As planned, we conducted a mixed within-between ANOVA on the agreement ratings with the attribution target (self vs. other), behavior type (action vs. belief), and their interaction as the predictor variables. Overall, there was a main effect of behavior type, such that subjects agreed that they and others could choose to keep the $20 ($M$ = 5.01, SD = 2.21), more so than they could choose to believe that keeping the $20 was not wrong ($M$ = 3.72, SD = 2.21), $F(1, 500) = 167.41$, $p < 0.001$, $\eta_G^2 = 0.08$. There was no main effect of target, as subjects did not judge themselves ($M = 4.28$, $SD = 2.35$) to have more control overall than others ($M = 4.45$, $SD = 2.25$), $F(1, 500) = 1.03$, $p = 0.311$, $\eta_G^2 < 0.001$. However, as predicted, we observed a significant interaction such that there was a discrepancy between self and other judgments for beliefs but not actions, $F(1, 500) = 5.05$, $p = 0.025$, $\eta_G^2 = 0.003$; see Figure 3.4, below. Because we observed non-normality in the action conditions, we conducted a non-preregistered Kruskal-Wallis rank sum test analyzing self vs other action-belief difference ratings. This test also revealed a significant difference of attribution target, $\chi^2(df = 1) = 6.96$, $p = 0.008$.

Follow-up independent-samples t-tests revealed that subjects rated themselves less able to choose to believe that keeping the $20 is wrong ($M = 3.52$, $SD = 2.20$) than their close other ($M = 3.92$, $SD = 2.21$), $t(499.99) = -2.02$, $p = 0.044$, $d = -0.18$. But there was no difference between self ($M = 5.04$, $SD = 2.25$) and other ($M = 4.98$, $SD = 2.18$) when comparing actions, $t(499.26) = 0.26$, $p = 0.793$, $d = 0.02$.[13] A follow-up

[13] One possibility is that the smaller difference in the action condition relative to the belief condition is due to a ceiling effect in the action condition. Consistent with this, even though the average ratings for actions were approximately 5 on the scale, the majority of values (55.3%) were a 6 or 7 on the 1-7 scale (see Figure 3.4). However, while we cannot definitively rule this out, there are two reasons to suspect that it is not the most likely explanation. First, compared with our prior research (Chapter 1), the average control rating of 5 observed in the present study is unusually low for an intentional behavior. In other studies with the same dependent variable, we have found that people's responses tightly cluster around a mean of 6 on a 7-point scale when judging others' control over their intentional behaviors. Thus, people could have used higher parts of the scale, and often do.

Another method for diagnosing a possible ceiling (or floor) effect is to examine the cumulative response distributions in order to determine whether there is divergence between conditions for values further from the ceiling/floor (see e.g., Simonsohn, Simmons, & Nelson, 2014). When we investigated the cumulative response distributions in the action condition, we observed a uniform lack of differences along every part of the scale. This suggests that the observed interaction is not the product of a ceiling effect in the action condition.

equivalence test showed that if there is a difference in attributions of capacity to commit the immoral act, it is smaller than *d* = .20, *t*(499.22) = 1.937, *p* = 0.027.

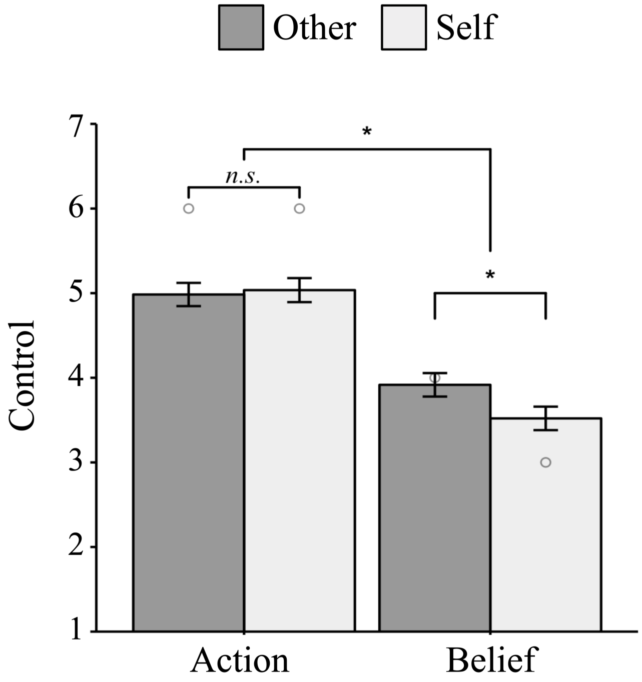

Figure 3.4. Means (and standard errors) of agreement ratings across conditions in Study 3.3. Circles represent median responses for each condition.

Discussion

Subjects reported that they were less able than a close other to change a belief that an immoral behavior was wrong, yet they reported no difference in their respective capacities to carry out that very same immoral behavior (ratings of action capacity were generally high for both self and other). These findings cast doubt on the idea that general self-presentational or self-enhancement concerns underlie the self-other asymmetry observed in Studies 3.1 and 3.2. Rather, they suggest that the true causal mechanism is isolated to judgments about beliefs (or mental states that are similar to beliefs).

A strength of Study 3.3 is that the contents of the belief and the corresponding behavior were closely matched. However, separate from content, the respective valences of these two stimuli also need to be considered – that is, how bad people thought it would be either to change the moral belief in question, or to carry out the described behavior. If subjects thought that changing their beliefs was worse than carrying out the corresponding behavior, then this might explain their reluctance to indicate a capacity for such belief change; and an alternative explanation based in self-presentation would thereby gain credibility. To check this possibility, we carried out three post-tests that compared people's moral judgments of the act described in Study 3.3 (picking up money that someone dropped and keeping it rather than returning it) with their moral judgments of holding the corresponding belief (belief that such an act is not wrong).

The three post-tests all called for subjects to make moral character judgments. In each case, subjects were initially presented with a description of the target behavior and indicated whether they agreed or disagreed that it was wrong. Only those subjects who agreed that the behavior was wrong continued with the remainder of the study. Post-test 1 compared performing an immoral act with judging (privately) that the act is not immoral. Post-test 2 compared performing the immoral act with professing (publicly) that the act is not immoral. Post-test 3 compared professing that one could not perform the immoral act with professing that one could not choose to believe that the act is immoral. Below is brief description of these studies; see Appendix L for full text of the vignettes.

Post-Test 1. In our first test, we compared subjects' impressions of someone who acts immorally with their impressions of someone who believes that the very same immoral act is not wrong. 56 subjects (out of 61) reported that the target behavior was morally wrong. These subjects then read a story involving three characters, Jones, Smith, and Peters. In the story, Jones accidentally and unknowingly dropped a $20 bill on the sidewalk. Smith, who was walking a little way behind him and saw this happen, picked up the $20 and pocketed it. Peters, who witnessed this scene from across the street, privately judged Smith's action as not wrong. Subjects reported their impression of each person. They judged Smith, who kept the $20 (M = -1.53 SD = 1.60) more negatively than Peters, who merely thought that this act was not wrong (M = -0.15, SD = 1.84), *t*(55) = -5.60, *p* < 0.001.

Post-Test 2. Our second test compared subjects' impressions of someone who publicly states that a wrong act is permissible with their impressions of someone who performs that same act. 59 subjects (out of 63) initially reported that the target behavior was morally wrong. These subjects then read a vignette in which they imagined themselves in conversation with the third party (Peters) in post-test 1. After witnessing Smith take the money, the subject turns to Peters and describes what they witnessed. Peters responds that he does not think the act was wrong (see Appendix L for full text). Subjects then made judgments of both Smith, who took the $20, and Peters, who proclaimed this act as not wrong. Subjects rated Smith (M = -1.88 SD = 1.57) slightly more negatively than Peters (M = -1.75, SD = 1.28), but the difference was not significant, *t*(58) = -0.83, *p* = 0.409. Thus, in both of post-tests 1 and 2, performing an immoral act was judged more negatively or no differently than believing or proclaiming that the same immoral act is in fact not wrong at all.

Post-Test 3. Our third test was more complex. It compared subjects' impressions of someone who publicly states that he could not perform an immoral act, with their impressions of someone who publicly states that he could not choose to believe that the same act is not wrong. This comparison was designed to alleviate a concern that subjects may report a lack of control over immoral beliefs because there is a distinctive reputational advantage from doing so, when compared with stating that one could not perform an immoral behavior. The claim that one could perform an immoral action may seem universally and uncontroversially true, particularly if it is interpreted in physical and not psychological terms (i.e., it clearly is physically possible to perform the act). It may therefore seem less diagnostic of a person's character than the claim that one could not choose to hold an immoral belief – which seems less universally true, more tied to idiosyncratic psychological factors, and therefore more diagnostic of a person's underlying character.[14]

To perform a complete test of this alternative explanation, we also asked subjects to judge two additional individuals – one who states that he could perform the act (though would not), and one who states that he could choose to believe that the act is not wrong (even though he does not believe this currently). Judgments of these individuals comprised a "baseline" against which to compare the original "could not" judgments, allowing a test of the relative "boosts" gained by denying the ability to act or believe immorally, respectively.

109 subjects (out of 118) initially reported that the target behavior was morally wrong and continued to the remainder of the study. They read the same basic vignette as

[14] We thank Josh Lewis and Joe Simmons for suggesting this possibility.

in the previous studies, describing Jones accidentally dropping $20 and Smith picking it up and keeping it for himself (there was no Peters in this vignette). Subjects were asked to imagine that they were a witness to this event, seated at a nearby table with several other people. The group then engages in a conversation about the event. The first critical comparison showed that a person who states to the rest of the group that, if he were in a similar situation, he could not choose to keep the money (M = 2.14, SD = 1.61), was judged more positively than a person who states that he could not choose to believe that taking the money is not wrong (M = 0.96, SD = 1.65), $t(107) = 6.11$, $p < 0.001$. We then compared both of these means against the two "baseline" conditions, in which the target individual stated that he could perform the act (M = -0.02, SD = 1.43), or could choose to hold the immoral belief (M = 0.00, SD = 1.49). These comparisons showed that the relative "boost" that an individual gains from asserting the lack of ability to hold an immoral belief (M = -0.96, SD = 2.24) is no greater, and is in fact significantly smaller than the boost gained from asserting the lack of ability to perform an immoral action (M = - 2.16, SD = 2.11), $t(107) = 5.16$, $p < 0.001$.

In sum, these three post-tests indicate the following: When an immoral act and an immoral belief are matched in content, (1) merely holding the belief is seen as less wrong than performing the act, (2) professing the belief is seen as no more wrong than performing the act, and (3) professing an inability to hold the belief is judged no more favorably than professing the inability to perform the act, and provides no more of a boost in impressions (when compared against the denial of these abilities). Indeed, if anything, there appears to be a stronger reputational advantage gained by claiming that one could not perform the relevant immoral behavior. On balance then, these post-tests indicate that there is no special reputational advantage that accrues from denying the ability to choose to hold immoral beliefs. Accordingly, the main finding from Study 3.3 – less control attributed to the self than to others over holding immoral beliefs, but no difference for immoral actions – seems unlikely to have resulted from self-enhancement or impression management motivations.

We argue instead that the results are explained by people's greater reliance on introspective experience when judging their own belief control as opposed to others'. When people consider their own ability to change a particular belief, they introspect on that belief and, in doing so, confront a feeling of low control. This feeling of low control drives down people's judgments of their own control over specific beliefs. However, this appreciation of low control is not generalized to others. What then, does account for how people judge others' control over their beliefs?

So far, we have said little about this process, and so we elaborate upon it here. By default, we postulate that people conceptualize beliefs as voluntarily controllable.[15] Accordingly, when people judge others' control over their beliefs, they apply this default, high control judgment to them – without introspection, and without careful consideration of the concrete specifics of those others' beliefs. This, of course, contrasts with the more

[15] We conducted a pilot study that lends further support to this claim. 267 University students from an elite college on the East coast were asked to rate their agreement (1: strongly disagree, 7: strongly agree) with four statements about how generally controllable beliefs are. Subjects indicated strong agreement with all four statements, including (1) "People can decide to believe something even when they have good reasons to believe the opposite." (M = 5.32, SD = 1.22); (2) "If someone really wants to believe that something is true, they can choose to believe it." (M = 5.35, SD = 1.28); (3) "People can make themselves believe whatever they want to." (M = 5.07, SD = 1.49); and (4) "No matter what, people can voluntarily choose to believe something if it benefits them to do so." (M = 5.22, SD = 1.42). These ratings were all significantly above the mid-point of the scales (*t*s > 11.2, *p*s < 0.001).

concrete consideration that is recruited when people consider their own specific beliefs, which evokes feelings of constraint. A natural consequence of these discrepant processes is the consistent self-other discrepancy observed in Studies 3.1-3.3.

By extension, we predict that people may also fail to incorporate their experience of low belief control into their abstract theory of belief change – not only as it applies to others, but as it applies to themselves as well. Thus, when people rate their own belief control in general, their judgments should resemble the judgments they tend to make about others. In Studies 3.4 and 3.5, we test two further predictions that follow from this reasoning. First, people should generally rate their own voluntary control over their beliefs as lower when they are considering a specific belief than when they are considering their belief control "in general" (Study 3.4). Second, the self-other discrepancy observed in the earlier studies should only emerge when people consider specific beliefs; it should disappear when people judge beliefs in general (Study 3.5).

Study 3.4

The goal of Study 3.4 was to test whether people more strongly agree that they have control over their beliefs when they judge control in general, as compared with when they judge their control over specific beliefs that they hold. Accordingly, we asked one group of subjects to indicate how much control they have over their beliefs "in general," and a separate group of subjects to rate their control over a set of specific beliefs that they currently hold. We hypothesized that the former (general) group would rely on their lay conceptualization of beliefs to answer this question, thereby reporting relatively high ratings of belief control. By contrast, we hypothesized that those asked to make judgments about specific beliefs would introspect on their beliefs and, in doing so, confront the practical limits on belief change, causing them to make lower judgments of control.

Method

Participants. 302 people (mean age = 37.7; 172 reported female, 2 unreported) were recruited from Amazon's Mechanical Turk (AMT) to participate in this study.

Design and Procedure. Subjects were told that we were conducting a study about people's general assessments of their life. At the start of the study, subjects were randomly assigned either to judge the degree of control they have over their beliefs in general (beliefs in general condition) or to judge their control over a set of specific beliefs they reported (specific beliefs condition).

In the "beliefs in general" condition, subjects rated their agreement (1 = *completely disagree*; 7 = *completely agree*) with a series of statements probing to what extent they thought they had control over different parts of their life, including where they live, their habits, and their job. Embedded in this set of statements was the target statement, "*My current beliefs are ones that I voluntarily hold. Specifically, I could change what I believe if I wanted to even if this means I was being wrong or immoral by doing so.*" In order to ensure attention, and to reduce acquiescence bias (given that control is generally seen as positive), all the statements were framed so as to highlight a possible downside of exercising control. For instance, in the work question subjects saw the statement "*My current job is one that I voluntarily hold. Specifically, I could change where I work if I wanted to even if this meant having a worse job.*" There were five statements in total, including the belief statement (see Appendix M). The statements were each shown on a separate page and in a random order.

In the specific beliefs condition, subjects wrote down a series of beliefs they currently hold and then, for each one, indicated whether they voluntarily held that belief. This happened in three stages. In the first stage, subjects were instructed to respond to the unconstrained prompt, “I believe that…” with the first belief that came to their mind. In the second stage, subjects responded similarly to a series of prompts intended to solicit beliefs about specific topics. These included beliefs about the subject’s work (“I believe that my work…”), their family (“I believe that my family…”), and themselves (“I believe that I…”), as well as a moral belief (“I believe that it is wrong to…”). These prompts were shown on separate pages and in a random order. In the third stage, subjects in the specific beliefs condition then rated how voluntarily controllable each of the earlier beliefs they produced was. For instance, if a subject filled in the first prompt with, “most people are good”, then they would respond to the following statement (bolding in the original):

> *Earlier you wrote that you **believe that… most people are good**. Please indicate whether you agree or disagree with the following statement: **I voluntarily hold this belief. If I wanted to, I could choose to <u>not</u> believe that most people are good**.*

Subjects indicated their agreement on seven-point rating scales (1 = *completely disagree*; 7 = *completely agree*) for each of the five concrete beliefs they had earlier produced. Each statement was shown on a separate page and in a random order.

At the end of the study, subjects from both conditions filled out a demographics form that asked them for their sex and age, and were then debriefed.

Results

Subjects assigned to the abstract condition generally agreed with the statement that, *in general*, they have voluntary control over their beliefs (*M* = 4.99, *SD* = 1.92). As predicted, this was significantly higher than subjects' average voluntary control ratings in the concrete condition (*M* = 3.96, *SD* = 1.44), *t*(299.96) = 5.34, *p* < 0.001, 95% CI [0.65, 1.41], *d* = 0.60. Furthermore, each specific belief prompt in the concrete condition yielded beliefs that, on average, were rated as less voluntarily controllable than ratings in the abstract condition. Beliefs generated by the unconstrained belief prompt were rated less voluntarily controllable (*M* = 4.45, *SD* = 2.11), *t*(263.59) = 2.32, *p* = 0.021, 95% CI [0.08, 1.01], *d* = 0.27, as were beliefs about work (*M* = 4.38, *SD* = 1.87), *t*(281.67) = 2.80, *p* = 0.005, 95% CI [0.18, 1.05], *d* = 0.32, the subject themselves (*M* = 4.13, *SD* = 2.10), *t*(264.07) = 3.66, *p* < 0.001, 95% CI [0.40, 1.33], *d* = 0.43, their family (*M* = 3.49, *SD* = 2.14), *t*(260.97) = 6.30, *p* < 0.001, 95% CI [1.03, 1.97], *d* = 0.74, and finally, morality (*M* = 3.36, *SD* = 2.13), *t*(261.89) = 6.87, *p* < 0.001, 95% CI [1.17, 2.10], *d* = 0.81.

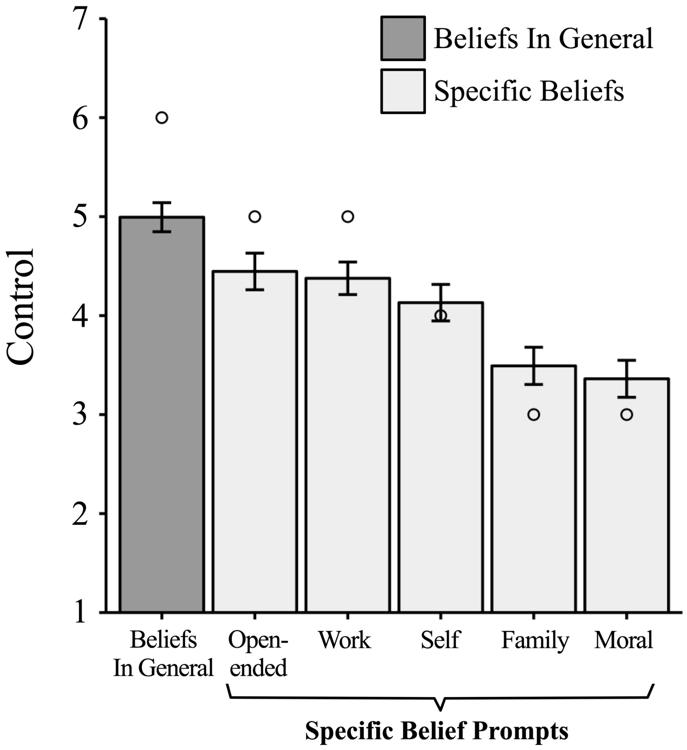

Figure 3.5. Means (and standard errors) of subjects' agreement with statements that they could choose to change their beliefs in general (dark bar), or that they could choose to change a range of specific beliefs (light bars). Circles represent median values.

Discussion

These results provide support for the hypothesis that people's general concept of belief controllability fails to match their specific experiences of belief control. When asked generally whether they have control over what they currently believe, subjects tended to indicate that they do have such control. This result is consistent with both the high ratings of control attributed towards others in Studies 3.1 and 3.2, as well as our pilot study (see Footnote 14). Yet, when asked about specific beliefs that they held, subjects attributed to themselves a lower degree of control. We observed this difference across every belief category we assessed, including beliefs about the subjects' themselves, their lives, and their morality. Most strikingly, even when subjects considered the very first concrete belief that came their mind, with no constraints on its content, they attributed lower control to themselves over this specific belief than when they considered their beliefs in general; thereby suggesting that it was not the specific belief prompts we used that gave rise to the observed differences.

In Study 3.5, we tested a final prediction that combines the self-other and concrete-abstract differences observed in the earlier studies. Our theoretical account explains these two phenomena in the same way. When people consider the amount of control they have over concrete beliefs that they hold, they are confronted with the psychological constraints on such control, which drives down judgments of their own belief control. But, when people attribute control to others (either concretely or abstractly), or when they attribute control to themselves in the abstract (without considering their specific beliefs), they rely on a more general theory according to which beliefs are quite controllable. Accordingly, both the self-other discrepancy and the concrete-abstract discrepancy arise because of a failure to integrate the experience of trying to control or change one's own beliefs with a more general model of belief controllability.

If this reasoning is correct, then the self-other asymmetry we observed in Studies 3.1-3.3 should be most pronounced when people are asked to consider specific beliefs, and should be attenuated or eliminated when people consider beliefs in general. We tested this prediction in Study 3.5.

Study 3.5

Method

Participants. We recruited 597 subjects (mean age = 35; 365 reported female, 231 reported male, 1 unreported) from Amazon's Mechanical Turk.

Design and Procedure. This study used a 2x2 between-subjects design. Subjects were randomly assigned to one of four conditions created by crossing belief condition (specific vs. general) and target condition (self vs. other).

At the beginning of the study, subjects were asked to think of a person close to them, write down that person's initials in a text box, and indicate their relationship to the person (they could indicate best friend, close family member, romantic partner, or spouse, or write in how they would describe the relationship). For ease of exposition, we will refer to this person as the "Close Other," or CO.

On the next page, subjects in all conditions wrote out four beliefs that they shared with the CO. They did so by completing four sentence fragments of the form, "We both believe that…". Subjects were instructed that the beliefs could be about anything, but had to be the first ones that came to mind.

In the specific-belief conditions, subjects were then presented with each of the four beliefs they had just written down (which they shared with their CO). For each belief, they reported their agreement with statements claiming that they or their CO had voluntary control over the belief. The specific contents of each belief were dynamically inserted within the corresponding statement. For instance, in the Self condition, the subjects would see the following:

> “You wrote that you believe that … [BELIEF]. Please indicate whether you agree or disagree with the following statement: I voluntarily hold this belief. If I wanted to, I could choose to not believe that [BELIEF].”

In the other condition, the statements were the same except that the initials of the CO were inserted into the statement:

> “You wrote that [INITIALS] believes that … [BELIEF]. Please indicate whether you agree or disagree with the following statement: [INITIALS] voluntarily holds this belief. If [INITIALS] wanted to, [INITIALS] could choose to not believe that [BELIEF].”

The order of presentation of the four beliefs was randomly determined for each participant.

In the general control conditions, subjects did not return to the specific beliefs they had reported. Instead, they indicated their agreement with four general statements about either their own, or the CO’s life, similar to those used in Study 3.4 (see Appendix N for full text). Only one of these statements – the belief statement – was relevant to our interests. The other statements were distractors, included simply to match the Specific condition in length. The key belief statement corresponded closely to the wording of the statements in the specific-belief conditions. In the self condition, subjects indicated their agreement with the statement, “My current beliefs are ones that I voluntarily hold. Specifically, I could choose to hold different beliefs if I wanted to even if this meant being wrong or immoral.” In the other condition, the initials of the CO were dynamically inserted into the statement as follows, “[INITIALS]’s current beliefs are ones that he/she voluntarily holds. Specifically, [INITIALS] could choose to hold different beliefs if he/she wanted to even if this meant being wrong or immoral." The distractor statements were about the subject's (or CO's) behavior, work, and home (e.g., "My current home is one that I voluntarily live in. Specifically, I could change where I live if I wanted to even if it meant changing many other parts of my life."). The order of the four statements was randomly determined for each subject.

All ratings were made on a seven-point rating scale (1 = "completely disagree"; 7 = "completely agree"). At the end of the study, all subjects reported their age and sex, and then were debriefed and paid.

Results

Examples of subjects' shared beliefs are provided in Table 3.2, below. As planned, we conducted an ANOVA predicting agreement ratings from attribution target (self vs. other), belief condition (general vs. specific), and their interaction. We observed a main effect of target such that subjects rated their own control ($M = 4.86$, $SD = 2.04$) lower than that of others ($M = 5.19$, $SD = 1.85$), $F(1, 593) = 4.27$, $p = 0.039$, $\eta_G^2 = 0.07$. We also observed a main effect of belief condition such that control ratings for specific beliefs ($M = 4.78$, $SD = 2.01$) were lower than control ratings for beliefs in general ($M = 5.27$, $SD = 1.86$), $F(1, 593) = 9.84$, $p = 0.002$, $\eta_G^2 = 0.016$. And, as predicted, these effects were qualified by a significant interaction, $F(1, 593) = 4.67$, $p = 0.031$, $\eta_G^2 = 0.008$ (see Figure 3.6).

Table 3.2
Examples of beliefs submitted by subjects in Study 3.5.

	Beliefs shared with close other			
Subject	**1**	**2**	**3**	**4**
A	Lifting is healthy.	Video games are fun.	Money is an important measure of success.	Intelligence is important.
B	Science is important.	Racism is bad.	Free healthcare should be a human right.	The world is doomed.
C	We are soulmates.	Abortion is wrong.	Our kids come first no matter what.	We have to make changes to improve our life.
D	We live comfortably.	We have the best children.	Parenting is tiring.	We have a bright future.
E	Cannabis is more-or-less harmless.	Social norms are too repressive.	Skateboarding is the best sport.	Video games are great fun.
F	There is a God.	Fat food makes us feel bad.	Our son is amazing.	A smaller house is better.
G	Cheating is bad.	Religion is dumb.	[Person] is dumb.	The earth is round.
H	Paul George will go to the Lakers.	Lebron is the best basketball player in the league.	“Impractical Jokers” is hilarious.	Going to the gym has benefits. beyond enhancing physical appearance.

Note. Subjects were instructed to write about four beliefs they shared with a close other that they had nominated on the previous screen. Each belief was elicited with the sentence fragment “We believe that…”.

As predicted, in the specific beliefs condition, subjects’ ratings of their own control ($M = 4.44$, $SD = 2.08$) were significantly lower than their ratings of others’ control ($M = 5.11$, $SD = 1.90$), $t(292.4) = 2.89$, $p = 0.004$, $d = 0.34$. However, there was no such self-other difference in the general belief condition (self, $M = 5.28$, $SD = 1.92$; other, $M = 5.27$, $SD = 1.80$, $t(296.68) = 0.062$, $p = 0.95$, $d < 0.01$. A follow-up equivalence test showed that, if there was a difference in the general belief condition, it is smaller than *d* = .2, *t*(296.77) = 1.69, *p* = 0.047. Examined another way, subjects rated their control over their nominated specific beliefs (*M* = 4.44, *SD* = 2.08) significantly lower than they rated their control over beliefs in general (*M* = 5.28, *SD* = 1.92), *t*(293.61) = 3.61, *p* < 0.001, *d* = 0.42, but there was no corresponding difference between ratings of others' specific (*M* = 5.11, *SD* = 1.90) and general (*M* = 5.27, *SD* = 1.80) belief control, *t*(295.83) = 0.72, *p* = 0.472, *d* = 0.08.

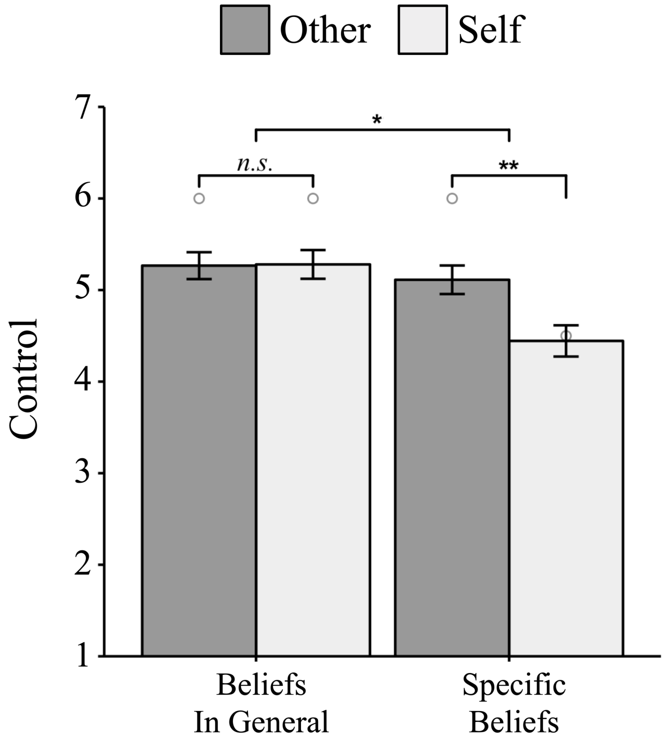

Figure 3.6. Means (and standard errors) for subjects' agreement ratings in Study 3.5. Circles represent median values.

Discussion

Results from Study 3.5 replicated findings from Studies 3.1-3.3 showing that, when considering control over specific beliefs, subjects judged themselves to have less control than others. As in Study 3.3, we observed this even when subjects rated beliefs they shared with a close other. Going beyond the earlier studies, we observed this discrepancy even though subjects were no longer constrained to consider a specific belief supplied by the experimenter, but were able to consider whichever specific beliefs came to mind. We also replicated Study 3.4's finding that people judge themselves as having less control when considering specific beliefs than when considering beliefs in general. This finding is noteworthy given that subjects provided four specific beliefs they held before indicating their control over beliefs in general – thereby showing that people maintain a general impression of belief control unless asked very directly about their control over specific beliefs. Most importantly, however, we found that the asymmetry between self and other control ratings was completely attenuated in the general beliefs condition – people do not judge themselves differently than a close other when it comes to a general assessment of belief control. These findings are consistent the theory we have articulated; namely, that self-other differences in belief control arise from the experience of psychological constraints on one's own ability to control specific beliefs, and from the subsequent failure to generalize this experiential understanding to others (and to integrate it within one's own theory of belief control).

General Discussion

Although control is a ubiquitous and important judgment in social cognition, determining behaviors such as blame and credit, and reward and punishment, it has not been extensively explored in relation to people's beliefs. In this paper we investigated how much control people attribute to themselves as compared to others. A great deal of past work suggests that people will inflate self-directed attributions of doxastic control. After all, control is highly desirable, and attributions of one's control are sometimes inflated as a consequence (e.g., Burger, 1986; Miller & Nelson, 1975). Recent research has even indicated that people tend to attribute to themselves greater free will over their own actions than they attribute to others, consistent with well-established self-enhancement biases (Pronin & Kugler, 2010). Despite these findings, we hypothesized that things would be different in the realm of beliefs.

A starting premise is that people are constrained in their ability to pick and choose what they believe. Specifically, when people perceive good reasons to believe something, these reasons limit their ability to choose to believe otherwise (e.g., James, 1937, see Introduction). However, famously, we also know that people often fail to appreciate the constraints that other people operate under, and that this failure can be especially pronounced for psychological constraints (Gilbert & Malone, 1995). That is, people often fail to appreciate that others' behavior is influenced by psychological constraints such as emotion or stress, which is representative of a broader failure to appreciate the complexity of others' inner lives (Johnson, 1987; McFarland & Miller, 1990; Miller & McFarland, 1987; Pronin, 2008). Accordingly, when it comes to belief change, people may also fail to appreciate the psychological constraints operating on others. This line of reasoning therefore generates a contrasting prediction, namely that people will tend to attribute to themselves less control over their beliefs than they attribute to others, at least insofar as they are considering concrete instances of belief.

We conducted a series of studies to adjudicate between these two possibilities. These studies consistently demonstrated that, when considering concrete instances of belief, subjects tended to attribute less control to themselves than to others. This self-other discrepancy occurred both when subjects held different beliefs from this other person (Study 3.1), as well as when they each held identical beliefs (Studies 3.2, 3.3, and 3.5). It arose not only when people considered strangers (Studies 3.1 and 3.2), but also when they considered close others (Studies 3.3 and 3.5). And it arose not only for beliefs supplied by us as the experimenters (Studies 3.1-3.3) but also for beliefs that subjects themselves supplied (Study 3.5). Thus, it appears to be robust.

In Studies 3.1 and 3.2, the self-other discrepancy emerged for judgments regarding the ability to voluntarily change beliefs, but not for judgments regarding whether those beliefs were intentionally chosen in the first place. At first blush, this result may seem out of step with our theorizing that feelings of constraint drive down self-attributions of control relative to others. And indeed, the difference across these two measures was not one that we had initially predicted. However, in hindsight, the lack of effect on the choice measure can be reconciled with the similar lack of a self-other difference when beliefs were considered abstractly. When an individual considers whether a particular belief was "intentionally chosen," he or she must make a retrospective judgment about an episode of belief formation in the past. Arguably, under these circumstances, the individual is psychologically distant from the actual experience of trying to control a belief. Similarly, when individuals judge the controllability of their beliefs in general, they are also psychologically removed from any immediate effort to control their beliefs, and therefore, do not feel the tug of low control that would typically arise if they had a specific belief in mind. In both cases, this psychological distance may prompt people to default to a theory of belief which posits high control, without directly considering what it would actually be like to control their beliefs. Thus, psychological distance from specific beliefs, however it arises, tends to equalize judgments of self and other, whereas psychological proximity tends to widen this gap.

These findings raise and important question; namely, why is it that people tend by default to regard beliefs as quite controllable, shifting away from this only when confronted directly with the experience of trying to change their beliefs? In other words, why would people hold an intuitive theory of belief that diverges from their experience (and possibly the reality) of belief? Precisely where this general theory of belief control comes from is uncertain. One possibility is that the general desirability of having control leads people unreflectively to attribute high control to beliefs. Another possibility is that people conflate control with the capacity to be rational, such that they intuitively associate the process of weighing reasons in order to form beliefs with the ability to exert voluntary control over those beliefs. Then, because they assume that people (including themselves) are generally rational when forming beliefs, they unreflectively conclude that people also have voluntary control over their beliefs. Lastly, perhaps people recognize the phenomenon of motivated reasoning (i.e., they recognize the influence that desires can have over beliefs), and mistake this phenomenon as providing evidence for voluntary control (see discussion below on this topic). The present results cannot locate the source of this general belief, but they suggest that it is somewhat unreflective, and can be unseated once people more deeply consider what is involved in controlling their beliefs.

Alternative explanations

One alternative explanation for the observed self-other discrepancy is that it reflects a form of self-enhancement (or impression management). Voluntarily controlling a belief merely because one wants to may be seen as violating or disrespecting the norms of belief (Clifford, 1877; James, 1937; Stahl & van Prooijen, J.W., 2018; Ståhl, Zaal, & Skitka, 2016). Accordingly, because of the negative connotations associated with this ability, subjects may be motivated either (a) not to perceive it in themselves (self-enhancement), or (b) not to report it to an experimenter, even when they do perceive themselves as having it (impression management). However, if either one of these alternatives explained the tendency to attribute less volitional control to the self than to others, then we should have observed a similar discrepancy for judgments about the ability to *act* poorly, since the ability to act in norm violating ways has similarly negative connotations. Yet, we observed no such self-other difference in attributions of the ability to carry out immoral acts that were matched in content to the respective beliefs (Study 3.3). Thus, this alternative explanation is not well supported by the evidence.

However, perhaps a subtler form of this alternative explanation can account for the results. What if people believe they has less control over their beliefs than others because they consider themselves to be more *objective* than others? Indeed, prior work coming from the framework of naïve realism has shown that people often assume they are objective perceivers and believers of the world (Pronin, Lin et al., 2002; Pronin, Gilovich et al., 2004; Ross & Ward, 1996). A natural corollary of assuming that one is objective, or that one perceives and judges the world “directly,” is the idea that one’s beliefs are outside one’s control. After all, if your beliefs are dictated by the way the world actually is, and not by some extra input on your part, then what you believe must be “limited” to what is objectively true. Perhaps then, even when people share a belief with a close other, they hold a prior assumption that they are more objective than the other person, and this in turn explains the discrepant self-other judgments of belief control.

One reading of this alternative explanation is that people hold a generalized assumption that their own reasoning is more grounded in objective processes than are others’ (Ehrlinger, Gilovich, & Ross, 2005). This interpretation is incompatible with our findings. A default belief in one’s own superior objectivity should apply just as readily to specific beliefs as it does to belief change considered in the abstract. But, in that case, this account cannot explain the results of Study 3.5, which revealed a self-other difference only for specific beliefs, and not at all for beliefs considered in the abstract. We therefore reject this rendering of a naïve realist alternative on the basis of Study 3.5’s results.

However, perhaps the belief about one’s superior objectivity is triggered only by an encounter with a specific belief that one holds. That is, perhaps it is only when people come to consider one of their beliefs concretely that they reason about their own objectivity and, once they have judged that they are objective, and that being objective implies that they cannot voluntarily change their belief, they attribute to themselves low control. This account posits that self-attributed objectivity is the pivotal mediating factor in this chain of inference, with belief constraint merely being an inferential by-product of this starting assumption. This would explain our findings, and we cannot rule it out definitevly. However, we would make three observations about it, which to our minds diminish its plausibility. First, it is implausible that people only believe they are objective reasoners in concrete situations – indeed, related research shows that people's self-attribute objectivity to themselves in both specific and general contexts (see e.g., Ehrlinger et al, 2005). Second, this account is less parsimonious and less plausible than our proposal in that it posits more intermediate psychological steps en route to a judgment of the self's lower belief control. We propose, instead, that all that is needed is that people encounter existing psychological constraints on belief change, and thereby recognize limitations to their ability to change a specific belief, without any need for an additional abstract inference of one's own greater objectivity.

But most importantly, an explanation for our finding couched in the framework of naïve realism assumes that ordinary people believe that objectivity entails constraint. However, we are not sure that the people in our studies believe this. It may be that ordinary people in general think that it is possible to choose whether or not to be objective when reasoning in the first place. This which would imply that they often think of their own beliefs as both objective and controllable. Given that people often believe that they are objective (both concretely and in the abstract; Ehrlinger et al., 2005), some of our findings – in particular, self's belief control in the abstract (see Studies 3.4 and 3.5), as well as high agreement that they deliberately chose their belief (Studies 3.1 and 3.2) – suggest this may be the case. Such simultaneous judgments of objectivity and voluntary control (and choice) run counter to the basic idea underlying a naïve realist explanation of our data – namely that greater perceived objectivity necessitates greater perceived belief constraint, and therefore lowered control.

In sum, there are several factors that incline us away from interpreting the present findings solely within a naïve realist theoretical framework. While it may be that one source of the sense of constraint over one's beliefs stems from an assumption of the self's greater objectivity, this source alone seems insufficient – it cannot parsimoniously explain the attenuation of the self-other discrepancy for general beliefs, nor can it easily accommodate conjoint judgments of objectivity and choice. For this reason, we see the findings presented here as supplementing theories of objectivity and bias attribution, rather than being subsumed by them. That said, the nuanced relationship between judgments of belief control and attributions of objectivity or bias is an important area for future research.

The relationship between *experienced* and *actual* control

In order to better understand the psychological processes responsible for the discrepancy we have discovered, it is necessary to know how lay judgments of control correspond to *actual* control. However, while our findings provide strong evidence that people attribute to themselves less control over their beliefs than they attribute to others, our data do not speak to which judgments (self-directed or other-directed) are more accurate. Are believers, who attribute relatively less control to themselves, or observers, who attribute relatively more control to those believers, more accurate?

A dominant view among psychologists and philosophers is that people do not have direct control over what they believe (Alston, 1988; Epley & Gilovich, 2016; James, 1897; Wegner, 1994; Wilson & Brekke, 1994; Gilbert, 1991, 1993). According to these scholars, people cannot simply believe whatever they want to because beliefs are spontaneous, automatic responses to information, akin to perceptual processes. If people have any control over what they believe, the story often goes, it is highly indirect (e.g., exerted only by controlling the information one is exposed to; e.g., Alston, 1988; Epley & Gilovich, 2016), or extraordinarily rare (e.g., choosing what to believe is possible only under conditions of extreme uncertainty or ambivalence; James, 1897; Sloman, Fernbach, Hagmeyer, 2010; though see Steup, 2017, for a dissenting view).

However, there is a dearth of empirical work on whether people can exert voluntary control over their beliefs. In fact, as far as we know, there is no direct evidence testing whether people can exert deliberate, voluntary control over their beliefs. Considerable work on the phenomenon of motivated reasoning and self-deception indicates that people's desires and preferences can impact their beliefs, but this research does not speak directly to the question of voluntary control. In fact, the evidence from these studies is consistent with the view that people lack voluntary control over their beliefs (e.g., Klein & Kunda, 1992, Kruglanski & Webster, 1996; and Kunda, 1990, for a review). For instance, evidence to date suggests that desires influence beliefs indirectly and unconsciously – by influencing how incoming information is interpreted, which information is stored in memory, and which information is subsequently retrieved (Pronin, Lin, & Ross, 2002; Epley & Gilovich, 2016; see Kunda, 1990, for an extended discussion of this). When psychologists give people direct evidence of their biases, remove ambiguity, or prevent biased retrieval, motivated reasoning and self-deception all but disappear (e.g., Bar-Hillel & Budescu, 1995; Sloman et al., 2010). Moreover, when people are given strong evidence in favor of some conclusion, they often heed this evidence, even when the conclusion is upsetting or undesired (see Wood & Porter, 2016, for recent experimental evidence; see Petty & Cacioppo, 1986, for a review).

Most philosophical arguments for the putative uncontrollability of beliefs are grounded in self-reports of low control (see Introduction). Our findings suggest that these self-reports are only partially shared by lay people (as compared with philosophers). The moderate ratings of control in our studies suggest that lay people appear to attribute to themselves considerably more control than do philosophers, although they do still clearly perceive some constraints on their ability to exert voluntary belief control. However, we should be wary of the idea that the experience of control is diagnostic of actual control. Introspection can be a poor guide to our mental processes or capacities, and prior work shows that people are notoriously poor at reasoning about the origin and quality of their beliefs (Davison, 1983; Hauser et al., 2007; Nisbett & Wilson, 1977; Pronin et al, 2002). There is also reason to think that people are poor judges of what they do and do not have control over (e.g., Buehler, Griffin, & Ross, 1994; Koehler & Poon, 2006). Recent evidence suggests that morality constrains people’s sense of choice, such that choosing between moral options feels more constrained than choosing between non-moral options (Kouchaki, Smith, & Savani, 2018). Yet, people who judge that they “could never” harm someone at time 1, may end up doing so at time 2 when incentivized the right way. For instance, people who are sexually aroused indicate a greater willingness to engage in morally questionable behaviors than their unaroused counterparts predict (e.g., falsely telling a partner that they love them just to increase the chance of having sex with that person, Ariely & Loewenstein, 2006). For this reason, external observers may sometimes be better judges of actors than actors are of themselves (e.g., Bass & Yammarino, 1991; MacDonald & Ross, 1999; Risucci, Tortolani, & Ward, 1989). Perhaps the people around us, who see our epistemic foibles more clearly than we do, are better informed about our capacity to capriciously choose our beliefs than we are.

In sum, our finding that people experience their beliefs as partially outside of their voluntary control provides prima facie evidence that this is indeed the case. However, in our view, the question is far from settled, as one's internal sense of control is not necessarily diagnostic of one's actual capacity. The bottom line is that without direct information about people's voluntary abilities to change their beliefs, we cannot know whether or how people are erring in their judgments in the present studies.

Implications for belief-based conflict

One important implication of this work is that believers and observers will sometimes disagree about the extent to which each person can change what they believe. This entails that in cases of disagreement, the two parties may disagree not only about who is wrong, but also about who is even capable of changing their mind, which may further exacerbate their disagreement. If one party to a disagreement believes that they are incapable of adopting the other party's view, then they will not be motivated to compromise in their view. But, if the other party expects the first person to compromise, and believes that doing so is within that person's control, then the fact that the first person does not do so may lead to judgments that she is culpably intransigent, and therefore deserving of blame or punishment. And because the first person believes that she cannot simply choose to change her mind, he will resent being blamed or punished for not doing so.

Well into this research, we discovered a real-world example that illustrates elements of this speculative drama. On April 10th, 2018, Megan McArdle, a conservative columnist, published an article voicing a widely-held thesis in conservative circles: that liberals are biased against conservatives on the basis of their beliefs, holding derisive attitudes about them that are unfair in the same way that prejudice directed towards other minorities is unfair (McArdle, 2018). The next day, Hamilton Nolan published a rebuttal to McArdle, offering a defense of the treatment of conservatives at the hands of liberals (Nolan, 2018). Tellingly, Nolan's article was titled "Ideology is choice," and his central argument went as follows: "unlike race and gender and sexual persuasion, it [being conservative] is an intellectual choice. It can be changed at any time." These authors, much like the subjects in our studies, appear to conceptualize each other's agency over their own beliefs differently, which in turn appears to lead to very different perspectives on their respective moral responsibilities.

Limitations and Future Directions

Subjects in our studies were exclusively recruited through Amazon's Mechanical Turk. Although samples recruited from AMT are more representative of the U.S. than typical university student samples, individuals on AMT tend to be less religious, wealthier, and better educated than the average person in the United States (Paolacci & Gabriele, 2014). Additionally, our entire sample consisted of people living in the United States who, like other so-called WEIRD populations, are wealthier and better educated than most people in the world, and are predominately Christian (Heinrich, Heine & Norenzayan, 2010). Cross cultural work has revealed striking differences in how different groups think about individuals' agency. Of particular note, individuals in some non-U.S. cultures appear to attribute less agency to individuals than do individuals in the United States (e.g., Iyengar and Lepper, 1999; Kitayama et al., 2004; Miller, Das, & Chakravarthy, 2011; Morris, Nisbett & Peng, 1995; Savani et al., 2010; Specktor et al., 2004). For instance, compared to children in the United States, Nepalese children are more inclined to view some behaviors as constrained by social rules and therefore outside of their control, with this gap widening with age (Chernyak et al., 2013). In a similar vein, Indian adults appear to be less likely than U.S. adults to construe everyday behaviors as choices (Savani et al., 2010). Of clearest relevance to the present studies, some work suggests that Christians tend to attribute more control to others over deviant mental states (e.g., consciously entertaining thoughts of having an affair) than do Jews, thus showing evidence for cultural moderation with respect to mental states in particular (Cohen & Rozin, 2001). In light of this sort of evidence, we should not automatically assume that the results from our studies will replicate across different cultural or religious contexts.

Although we are uncertain as to whether our findings will generalize to all cultures, our findings do suggest an important direction for cross-cultural work. Specifically, future work measuring attributions of belief control should distinguish between lay theories of belief control and the introspective-experience of belief control. One virtue of measuring both is that we may expect different amounts of variation between these two measures of control across cultures. For instance, assuming that beliefs are indeed uncontrollable to a significant degree (see above), we should expect that the felt-experience of low control will vary little from culture to culture. By contrast, the lay theory of belief, which may be influenced by highly variable norms (e.g., religious norms, Cohen & Rozin, 2001), or folk theories of agency (see paragraph above), may be more likely to vary across cultures. For this reason, we speculate that self-other differences in belief control are most likely to arise in cultures where the lay theory of belief posits high control, as it is in these cultures where this lay theory will most likely diverge from the felt-experience of belief.

Another limitation in our studies regards the limited range of beliefs that we sampled. The beliefs in Studies 3.1-3.3 were highly abstract, complex, or value-laden (e.g., belief in God, the correct policy for genetically modified foods, the wrongness of not returning money to its rightful owner). We addressed this in Studies 3.4-3.5 by using beliefs that subjects themselves provided – specifically, the first beliefs that came to mind. This yielded a considerably wider sampling of belief contents (see Table 3.2 for a list of examples). Yet, it still leaves open the question of how people reason about their own control relative to that of others for very simple, concrete beliefs (e.g., "there is a two thirds chance of pulling a marble out of the bucket," "there is a quarter in my pocket," "it is raining"). We are ambivalent about whether to expect the same discrepancy in cases such as these. It may be that the self-other difference is attenuated or eliminated given that the relevant constraints on belief change are far more apparent for beliefs of this sort. Continuing to delimit the bounds of the self-other discrepancy remains a valuable goal for future research.

Finally, research should investigate whether, and when, self-other differences in attributions of belief control extend to other mental states. Although the present paper focuses only on the constraints on belief change, it may be that other mental states, including desires, evaluative attitudes, and emotions, are subject to similar constraints. If they are, then we might expect similar self-other discrepancies in perceived control – particularly in light of past work showing that people generally attribute high control to others over many mental states (Cusimano & Goodwin, in press). Indeed, there is already one reason to expect the self-other discrepancy to extend to other mental states, namely, that a person's beliefs often play a pivotal role in determining his or her other mental states. For instance, if someone is depressed because she believes she will not recover from a severe illness, an observer may think she is more capable of cheering up than she herself does, precisely because the observer judges her as more able to change her belief about her prognosis than she does. However, whether such self-other differences do in fact extend to other mental states awaits empirical testing.

Conclusion

The present paper uncovers an important discrepancy in how people think about their own and others' beliefs. Put succinctly, when someone says, "You can choose to believe in God, or you can choose not to believe in God," they may often mean that *you* can – *they* cannot. In other words, people judge that others have a greater capacity to voluntarily change their beliefs than they, themselves do. We argued that this derives from two distinct ways people reason about belief control: either by consulting their lay theory of belief change, or by introspecting and reporting what they feel when they consider voluntarily changing a belief. When people apply their lay theory of belief, they judge that they and others have considerable control over what they believe. But, when people consider the possibility of trying to change a particular belief, they tend to report that they have less control. Because people do not have access to the experiences of others, they rely on their lay theory of beliefs when judging others' control. Discrepant attributions of control for self and other emerge as a result. This may in turn have important downstream effects on people's behavior during disagreements. More work is needed to explore these downstream effects, as well as to understand how much control people actually have over what they believe. Predictably, we find the results from these studies compelling, but admit that readers may believe whatever they please.

APPENDICES

Appendix A

Behaviors and mental states used in Study 1.1

<u>Behavior Foils:</u>

accidental act
dropped
fell off of
ran into
slipped on
tripped over

uncontrollable act
fainted
shivered
sneezed
sweated
yawned

intentional act
ate
avoided
played with
said
searched for

<u>Mental States:</u>

intention
aimed to
decided to
determined to
had the goal to
intended to
meant to
planned to
plotted to
resolved to
willed to

deliberation
considered
contemplated
deliberated about
interpreted
pondered
rationalized
reasoned about
ruminated about
speculated about
thought about

imagination
imagined
pictured
pretended
visualized

belief
accepted that
assumed that
believed that
concluded that
decided that
expected that
feared that
felt that
figured that
guessed that
had faith that
had the impression that
intuited that
judged that
posited that
suspected that
thought that
trusted that
understood that
was confident that

evaluation
appreciated
disapproved of
disliked
enjoyed
hated
liked
loved
respected
revered
valued

desire
ached for
coveted
craved
desired
hoped for
longed for
lusted after
wanted
wished for
yearned for

emotion
felt afraid
felt amused
felt angry
felt anxious
felt depressed
felt disgusted
felt embarrassed
felt happy
felt irritated
felt sad

memory
forgot
recalled
recognized
remembered
repressed

Appendix B

Pairwise differences in behavior/mental state category in Study 1.1

We computed a series of paired t-tests between all categories for Control (Table B.1) and Intentionality (Table B.2). All p-values were Holm-Bonferroni-corrected within, but not between, control DV.

Table B.1

Dependent-sample t-test values between experimental conditions on ratings of control.

	1	2	3	4	5	6	7	8	9	10
1. Uncontrollable										
2. Accident	-9.35									
3. Emotion	-10.23	-2.02								
4. Memory	-12.38	-3.81	-1.91							
5. Desire	-13.64	-6.34	-4.75	-3.67						
6. Evaluation	-17.34	-10.42	-9.79	-9.53	-6.40					
7. Belief	-17.47	-11.24	-9.17	-9.48	-7.77	-0.72				
8. Imagination	-19.56	-13.81	-12.76	-13.25	-10.75	-6.19	-4.38			
9. Deliberation	-18.37	-13.18	-12.46	-11.31	-9.84	-6.08	-4.59	-0.84		
10. Intention	-23.60	-19.78	-17.85	-18.24	-16.01	-14.65	-12.99	-11.21	-9.11	
11. Intentional Act	-23.77	-20.68	-18.20	-18.20	-16.37	-15.20	-12.84	-12.73	-9.86	-2.03

Note: Underlined values are non-significant following Holm-Bonferroni correction. *df* = 142.

Table B.2

Dependent-sample t-test values between experimental conditions on ratings of intentionality.

	1	2	3	4	5	6	7	8	9	10
1. Uncontrollable										
2. Accident	-2.70									
3. Emotion	-10.37	-7.74								
4. Memory	-14.65	-11.03	-4.31							
5. Desire	-17.64	-15.19	-11.16	-8.77						
6. Evaluation	-20.95	-18.39	-14.88	-12.80	-5.15					
7. Belief	-21.18	-19.12	-13.65	-13.63	-6.87	-1.54				
8. Imagination	-23.40	-22.44	-18.00	-15.93	-10.12	-7.18	-4.44			
9. Deliberation	-20.93	-19.34	-15.56	-14.03	-9.40	-6.98	-4.97	-1.31		
10. Intention	-25.98	-25.15	-21.16	-19.93	-15.48	-15.22	-12.37	-12.09	-8.26	
11. Intentional Act	-26.71	-25.69	-21.64	-19.75	-15.05	-15.01	-12.03	-12.10	-8.11	-0.49

Note: Underlined values are non-significant following Holm-Bonferroni correction. *df* = 142.

Appendix C

Study 1.2 Stimuli

Paul is currently attending college where he studies English and Economics. His grades have been slipping and, because of this, his mother has been harassing him to study more. She eventually decides to visit him but, as she is driving up, ends up in a terrible car wreck. She is quickly found and rushed to a nearby hospital. Despite his mother's injuries, the doctors inform Paul that they expect she will survive.
Upon hearing this, Paul…
feels unhappy/angry that his mother might survive
wants/desires his mother not to survive
thinks/believes that life would be better if mother did not survive
dislikes/hates that his mother might survive.

James is a 50-year-old white male. He grew up in a middle-class family and is currently a manager at a bank. He married a few years after graduating college and he and his wife have a daughter. James's daughter is currently living and working in another state and has just called to tell her parents she has entered into a serious relationship. Over the course of the phone call it becomes clear that her boyfriend is African American.
When he hears this, James…
feels unhappy/angry that his daughter is dating an African American
wants/desires his daughter not to be dating an African American.
thinks/believes that it is wrong for his daughter to be dating an African American.
dislikes/hates that his daughter is dating an African American

Wesley is in his late 20s. He works at a used bookstore in his hometown and tries to stay informed about politics and current events. Recently he has been following one event in particular: a terrorist group had captured a city and was likely going to publicly execute dozens of dissidents in hiding there. In response, the United Nations (UN) has launched a counter attack which, according to analysts, was likely to succeed given the UN force's relatively superiority.
While watching all of this on the news, Wesley…
feels unhappy /angry that that the UN's counter attack will likely succeed
wants/desires the UN counter attack not to succeed
thinks/believes that it would be better if the UN counter attack did not succeed
dislikes/hates that the UN counter attack will likely succeed.

Amy is a college student writing about the use of torture for a political science senior thesis. Her thesis is about what methods of torture were typically successful or unsuccessful in breaking people's resistance. Digging through some archives, she found video footage of a journalist being beaten and tortured by secret police.
While watching the footage, Amy…
dislikes/ hates that the journalist wasn't in even more pain
wants/desires the journalist to be in even more pain

thinks believes that it would be better if the journalist was in even more pain
feels unhappy/angry that the journalist wasn't in even more pain

Appendix D

Study 1.3 Stimuli

Paul:

Non-constraining causal history:

Paul has a developmental disorder and, as a result, has difficulty speaking in fluid sentences.

Despite this disability, Paul has made it to college where he is currently studying Economics. However, lately, his grades have been slipping. Because of this, his mother, Vanessa, has been harassing him to study more. Her harassment has begun to be an inconvenience for Paul.

Vanessa eventually decides to visit her son but, as she is driving up, ends up in a terrible car wreck. She is quickly found and rushed to a nearby hospital. The doctor's call Paul to tell him about what happened to his mom. They are uncertain about what will happen to her.

Paul is still thinking about what a pain his mother has been lately and, in that moment, likes the idea of his mother passing away.

Low Control

Paul has a severe developmental disorder and, as a result of this disorder, lacks the ability to feel empathy for others or form normal familial bonds with them.

Despite this disability, Paul has made it to college where he is currently studying Economics. However, lately, his grades have been slipping. Because of this, his mother, Vanessa, has been harassing him to study more. Her harassment has begun to be an inconvenience for Paul.

Vanessa eventually decides to visit her son but, as she is driving up, ends up in a terrible car wreck. She is quickly found and rushed to a nearby hospital. The doctor's call Paul to tell him about what happened to his mom. They are uncertain about what will happen to her.

Paul is still thinking about what a pain his mother has been lately and, in that moment, likes the idea of his mother passing away.

James

Non-constraining causal history:

James is a 50-year-old white male. He grew up in a middle-class family and is currently a manager at a bank. He married a few years after graduating college and he and his wife have a daughter.

James's daughter is currently living and working in another state and has just called to tell her parents she has entered into a serious relationship.

Over the course of the phone call it becomes clear that her boyfriend is African American. Although James does not say anything to his daughter, he dislikes that his daughter is dating an African American.

James's father was an overbearing busy-body who tried controlling every aspect of his children's lives. All of James's siblings make judgments about their children life choices as well.

Constraining causal history:

James is a 50-year-old white male. He grew up in a middle-class family and is currently a manager at a bank. He married a few years after graduating college and he and his wife have a daughter.

James's daughter is currently living and working in another state and has just called to tell her parents she has entered into a serious relationship.

Over the course of the phone call it becomes clear that her boyfriend is African American. Although James does not say anything to his daughter, he dislikes that his daughter is dating an African American.

James's father was a hateful person who constantly told his children that black people were dangerous and irresponsible. All of James's siblings have attitudes like this deeply ingrained in them.

Wesley

Non-constraining causal history:

Wesley is in his late 20s. He works at a used bookstore in his hometown and tries to stay informed about politics and current events.

Recently he has been following one event in particular: a terrorist group had captured a city and was likely going to publicly execute dozens of dissidents in hiding there. In response, the United Nations (UN) has launched a counter attack which, according to analysts, was likely to succeed given the UN force's relatively superiority.

While watching all of this on the news, Wesley dislikes that the UN counter attack will likely succeed.

Wesley slipped in the shower and broke his arm about a year ago. While he is healthy again, his movement in his arm is still restricted and is occasionally sore. The doctors suspect that his muscles will never fully recover.

Constraining causal history:

Wesley is in his late 20s. He works at a used bookstore in his hometown and tries to stay informed about politics and current events.

Recently he has been following one event in particular: a terrorist group had captured a city and was likely going to publicly execute dozens of dissidents in hiding there. In response, the United Nations (UN) has launched a counter attack which, according to analysts, was likely to succeed given the UN force's relatively superiority.

While watching all of this on the news, Wesley dislikes that the UN counter attack will likely succeed.

Wesley slipped in the shower and hit his head about a year ago. While he is completely healthy again, his worldview has changed in a lot of ways. The doctors suspect that this is because his brain chemistry is different, which is affecting, among other things, his thoughts and beliefs.

Amy

Non-constraining causal history:

Amy is a college student writing about the use of torture for a political science senior thesis. Her thesis is about what methods of torture were typically successful or unsuccessful in breaking people's resistance.

Digging through some archives, she found video footage of a journalist being beaten and tortured by secret police. While watching the footage, Amy likes that the journalist is in a great deal of pain as this makes for a better senior thesis.

Amy is the first person in her family to go to college. Her father has been pressuring her to succeed ever since she was a child to the point where her entire identity has become about school. In addition to working on a senior thesis, her father has insisted that she take graduate-level coursework and run for student government.

Constraining causal history:

Amy is a college student writing about the use of torture for a political science senior thesis. Her thesis is about what methods of torture were typically successful or unsuccessful in breaking people's resistance.

Digging through some archives, she found video footage of a journalist being beaten and tortured by secret police. While watching the footage, Amy likes that the journalist is in a great deal of pain as this makes for a better senior thesis.

Amy is the first person in her family to go to college. Her father has been pressuring her to succeed ever since she was a child to the point where her entire identity has become about school. So, matter what she does, when she thinks about the journalist's pain, she is numb to it. Instead, her mind turns to the thought of failing her thesis, not graduating with honors, and disappointing her father.

Appendix E

Scales used in Study 2.1 Survey 2

Lay theories of Intelligence Scale

(1: strongly disagree, 5: strongly agree)

1. I don't think I personally can do much to increase my intelligence.
2. I believe I have the ability to change my basic intelligence level considerable over time.
3. Regardless of my current intelligence level, I think I have the capacity to change it quite a bit.
4. To be honest, I don't think I can really change how intelligent I am.

Life Satisfaction Scale

(1: strongly disagree, 5: strongly agree)

1. In most ways my life is close to my ideal.
2. The conditions of my life are excellent.
3. I am satisfied with my life.
4. So far I have gotten the important things I want in life.
5. If I could live my life over, I would change almost nothing.

Lay theories of emotion scale

(1: strongly disagree, 5: strongly agree)

1. If I want to, I can change the emotions that I have.
2. I can learn to control my emotions.
3. The truth is, I have very little control over my emotions.
4. No matter how hard I try, I can't really change the emotions that I have.

Appendix F

Study 2.1 Stimuli

Camping / Distress

You and your friend, Arthur, are on a hike in the woods outside a cabin you are renting with a group of friends. It is close to lunch time and you need to turn back to meet up with the others.

After a few minutes of hiking you hear a thud behind you followed by Arthur saying “Ow!”. When you turn around you see that Arthur tripped on a root and fell down. You walk to where Arthur is on the ground and ask if he is okay. As he sits up you can see he is wincing and breathing heavily, his hands are grabbing his leg.

He seems distressed, but you cannot see why. You do not see any scratches on his leg, nor any blood. It doesn’t even look like a bruise is forming. It looks to you like he is barely injured, if at all.

You help Arthur get to his feet so you can both continue on. But he immediately sits down again and says, “I can’t handle the pain – I need to sit a while before I can continue on.” You can see him clenching his teeth.

You are worried that if you two stay here you are going to be late to lunch with the rest of your friends. You two may also miss out on whatever else your friends had planned for the afternoon.

Vacation / Stress

You and a group of friends are vacationing in a beach house for a weekend. This getaway was scheduled a few months ago and you have been looking forward to it ever since. After arriving at the house, you go for a walk on the beach. One of your friends, Joe, joins you.

The weather is perfect, the sand is soft and warm, and it is close to sunset. But when you go to remark on this to your friend Joe, you notice that he is seems distracted, and not at all enjoying himself. When you ask him what is going on, he tells you, “I’m sorry. I’ve got this thing for my class on Monday and I’m feeling really stressed about it.”

The two of you turn around and start heading back to the house. Joe is quiet and looking down most of the way. You finally ask him what he is stressed about and he tells you, “It is my turn to read a poem in front of the class.” You ask him if he has to write the poem himself and he says, "No, we have all been picking one from a set the professor provided." You think to yourself that you have seen Joe talk in front of the class and he always does a good job.

Joe then adds, “Thanks for talking with me but I don't think I should have come on this walk – I’m too stressed. I don’t think I’ll feel relaxed until after the whole thing is over.”

You think to yourself that, if Joe is stressed the entire weekend, the trip will have been a waste.

Exam / Upset

You haven’t spoken to your friend, Rebecca, in a couple days. You two have plans later tonight to see some other friends. You call her to confirm them but, when you ask how she is doing, she responds that she is upset, and has actually spent the last couple hours in her room crying. When you ask what’s wrong she tells you that she just got her exam back and that she, in her words, “got a really bad grade on it”.

You recall that she took an exam last week in her Biology 101 course. You remember because you had invited her out for a drink but she said she needed to stay in because she had an exam the next day.

After a minute, you ask her how she actually did on the exam. She tells you that she got an A- on it.

It would be nice to see Rebecca later, but you also know that if she is still feeling upset it will bring everyone else down. You ask her if she is still planning on going out and she responds, between sniffles, “I don’t know. I’m still feeling upset.”

Thanksgiving / Embarrassed

You are visiting your family for the holidays. You are especially excited because aunts, cousins, and grandparents are all visiting, and this is your first opportunity in many years to see everyone together.

You have invited the person you are currently dating, Jamie. Jamie wants the two of you to cook something to impress the rest of the family and recommends a two-tier cake. While you are mixing it together some family members come in and tell you how excited they are to try it. You and Jamie put the cake in the oven, set the timer, and go off to chat with everyone.

25 minutes later you come back because you smell burning coming from the oven. You rush to take the cake out and… it is fine! You caught it right as the top started to singe but you were able to peel that off and cover it in icing. You tried it and it tastes just as you hoped it would. The cake looks a little funny on top but is otherwise fine.

When you look at the oven you see that the temperature was set to 375 instead of 350. When you point this out, Jamie turns a deep shade of red – they were the one who mistakenly set the temperature.

Your family is in the living room trading stories and having fun. Instead of joining them, Jamie heads upstairs to the room where you two are staying. You go up there and try to entice them to come down and join everyone, but Jamie says to you, “Don’t pressure me to join everyone! I’m so embarrassed about ruining the cake. I wanted to make a good impression and I messed it up!”

You want to go downstairs and spend time with the family you rarely see, but you do not want to leave Jamie all alone feeling bad, either.

Appendix G

Study 2.2 Vignettes

Note: Text that varied between condition is demarcated by brackets.

Grade

You haven't spoken to your friend, Ian, in a few days, but you two have plans later tonight to see each other. You call him to confirm but, when you ask how he is doing, he tells you that he is upset, and has actually spent the last hour in his room feeling sad. When you ask what's wrong, he tells you that he just got his exam back and that he, in his words, "got a really bad grade on it".

You recall that he took an exam last week in his Biology 101 course. You remember because you had invited him out for a drink, but he said he needed to stay and study for an exam the next day.

After a minute, you ask him how he actually did on the exam. He tells you that he got an [C-] [A-] on it.

It would be nice to see Ian later, but you also know that if he is still feeling upset it will bring everyone else down. You ask him if he is still planning on going out and he tells you, "I don't know. I'm still feeling upset."

Rain

There was a particularly bad storm last night where you live – the sound of the wind and rain kept you awake most of the night. The next morning you call your friend, Mike, to talk about it. It becomes immediately clear that he is deeply upset. He tells you that, last night in the storm, the wind pushed one of his windows open, letting the rain into part of his room.

You ask him if any damage had been done and he emphatically yells "Yes!". Apparently, the window was over his desk, so some unimportant papers and receipts are soaked and have to dry. [But, he adds, his collection of Sudoku books was also on his desk and has been completely destroyed by water damage.]
[But, he adds, his laptop and $1500 camera were also on his desk and both have been completely destroyed by water damage.]

As he describes this, it seems to you like he is on the verge of tears. He is going on and on about how much it is going to cost to replace everything, and how this has ruined his entire weekend.

Beer

You are out at a bar with your friend, Eric. The two of you are gossiping and having a good time. He wants to show you some photos he recently took and goes to pull his phone out of his jacket. However, his jacket is puffy and, when he pulls his phone out of his pocket, he accidentally knocks his beer over with his elbow.

[The beer spills on him, soaking his jeans and staining his favorite shirt. It does not spill on you or anyone else, but what was left of his beer is now gone.]
[The beer spills a little on the counter but does not get on him, you, or anyone else. What is left of his beer is now gone.]

You help him clean up, but it quickly becomes obvious that Eric's mood has soured. He is quiet and now looks quite upset. When you try to laugh it off and order a new beer, he says he doesn't want one. A few minutes later he tells you that he wants to go home, even though it is still quite early in the evening.

Birthday

You are getting dinner with your friend, Anthony. You two are chatting, including joking about others and airing your various grievances.

After a few minutes, though, Anthony admits to you that he's been upset all day because his sister really hurt his feelings.

When you ask him what she did, he tells you that she didn't call him last week to wish him a happy [Half-Birthday] [Birthday]. What's more, he adds, she still hasn't called.

Although you were joking around earlier in the conversation, Anthony now seems genuinely upset. His shoulders slump down and he looks away. He tells you that he has lost his appetite and falls silent.

Boyfriend/Girlfriend

You are talking to your friend, Andrew, on the phone. You two are catching up and he mentions that he is feeling distraught about his girlfriend, Olga.

[When you ask him what is going on, he tells you that, right at this moment, his girlfriend is getting coffee with a male colleague from her job. He says they are discussing how to deal with a serious problem at work, but that he is nevertheless feels upset about Olga getting coffee with another man.]
[When you ask him what is going on, he tells you that, right at this moment, his girlfriend is getting coffee with her ex fiancé. He says they are apparently just catching up, but that he nevertheless still feels upset about Olga getting coffee with him.]

As your conversation continues, Andrew keeps changing the topic to talk about what Olga is up to at that moment. When he talks, he furrows his brow and shakes his leg violently under the table.

Suicide

You are on a weekend beach vacation with your friends. You are trying to have fun and enjoy the time off but your friend, Jamie, is not having a good time. He has been despondent the entire trip: he isn't engaging people in conversation, laughing, or participating in group activities. Jamie's attitude is wearing on everyone else: you and your friends are walking on eggshells around him.

[When you and your friends asked him what is going on, he tells you that he is upset because he was talking to his father right before the trip and he told him that his mom is in the hospital again after a suicide attempt.]
[When you and your friends asked his what is going on, he tells you that he is upset because he was talking to his father right before the trip and he told him that the National Institute of Health has released a report saying that the US suicide rate has increased 1.5% in the past year.]

When Jamie stops talking, he shrinks into her seat and resumes staring off into the distance.

Appendix H

Study 2.3 Vignette

You are on a weekend beach vacation with your friends. You are trying to have fun and enjoy the time off but your friend, Jamie, is not having a good time. He has been despondent the entire trip: he isn’t engaging people in conversation, laughing, or participating in group activities. Jamie's attitude is wearing on everyone else: you and your friends are walking on eggshells around him.

> Calibrated Condition:
> When you and your friends asked him what is going on, he tells you that he is upset because he was talking to his father right before the trip and he told him that his mom is in the hospital again after a suicide attempt.
>
> Mis-calibrated Condition
> When you and your friends asked his what is going on, he tells you that he is upset because he was talking to his father right before the trip and he told him that the National Institute of Health has released a report saying that the US suicide rate has increased 1.5% in the past year.
>
> When Jamie stops talking, he shrinks into his seat and resumes staring off into the distance.

You know that if Jamie remains this upset, everyone's trip will be ruined.

Appendix I

Study 2.6 Vignettes

Shane:
Shane is an African American political science major at a high-ranking public university in the American Midwest.

Shane recently told his academic advisor that his college experience has been upsetting on account of his status as an African American, that he feels depressed, and that he is thinking of transferring schools.

When asked for examples, Shane responds, “Other students cross the street when they see me walking their direction, or they stare at me whenever I’m around them. During Halloween, I see students dress up in black face. Even though I work hard, this stuff makes me feel like I don’t belong here.”

Manuel:
Manuel is a freshman in a large public university in California. He is the first student from his family to go to college.

Manuel gets the impression that, as a student of Latin descent, his professors and other students seem to expect less of him in class compared to White students. He gets this impression from frequent patronizing comments in class. Outside of class, he is frequently mistaken for the janitor (or other university staff member) rather than a student.

Manuel now constantly feels anxious in class. And when other people mistakenly believe he is not a student he feels angry and upset. He is considering dropping out and moving back home.

Amy:
Amy is a Chinese American art history major at a private university in the North-east United States. When asked about her experience as a Chinese American student, she said that she has spent most of college feeling anxious about the way that other students and teachers treat her.

According Amy, professors frequently mistaken her for other Asian students in their classes. She is also often asked by others to answer a question from the perspective of someone who is Asian. She has told people that this makes her feel like she has nothing else to offer in class, but people continue to do it.

Because of this, she feels unmotivated to work and no longer feels comfortable contributing to class discussions. As a result of this, her grades have started to decline.

Appendix J

Belief statements used in Study 3.1.

Which of the following statements best reflects your views on the use of genetic modification in food production?
I believe that genetically modified foods should be prohibited.
I believe that genetically modified foods should not be prohibited.

Which of the following statements best reflects your views on how to best combat global climate change?
I believe that global climate change is a problem that is best addressed through strong government regulation.
I believe that global climate change is a problem that is not best addressed through strong government regulation.

Which of the following statements best reflects your views on the existence of God?
I believe that God exists.
I believe that God does not exist.

Which of the following statements best reflects your views on the impact that social media has had on dating?
I believe that social media has had an overall negative effect on dating.
I believe that social media has not had an overall negative effect on dating.

Appendix K

Belief statements used in Study 3.2

Genetically modified foods should be prohibited.
Global climate change is a problem that is best addressed through strong government regulation.
God exists.
Social media has had an overall negative effect on dating.

Appendix L

Full text of scenarios from post-tests associated with Study 3.3

Impression scale:

extremely negative			not negative or positive			extremely positive
-3	-2	-1	0	1	2	3

Text from Post-test 1:
Suppose you are walking down the street. Ahead of you are two people, **Jones** and **Smith**. Jones is walking about 20 paces ahead of Smith. They do not know each other. // Jones accidentally drops a $20 bill on the ground when he pulls his hand out of his pocket. He does not notice that he did this. // When Smith reaches the $20, he picks it up and puts it in his pocket and then changes direction, clearly intending to keep the money. // When you see this, you point it out to the stranger standing next to you, **Peters**. However, Peters says he doesn't think it was wrong what Smith did. // How much do these events impact your impression of each of these people?

My impression of Jones is…
My impression of Smith is…
My impression of Peters is…

Text from Post-test 2:
Two people, **Jones** and **Smith**, are walking down the street. Jones is walking about 20 paces ahead of Smith. Across the street is **Peters**, who can clearly see Jones and Smith. None of them know each other. // Jones accidentally drops a $20 bill on the ground when he pulls his hand out of his pocket. He does not notice that he did this. // When Smith reaches the $20, he picks it up, puts it in his pocket, and then changes direction, clearly intending to keep the money. // Across the street, Peters sees Smith keep the $20 instead of returning it and thinks to himself "What that guy did is not wrong". // How much do these events impact your impression of each of these people?

My impression of Jones is…
My impression of Smith is…
My impression of Peters is…

Text (and questions) from Post-test 3:
Suppose you sitting down to eat with some people and you are placed at a table with a clear view of the sidewalk. // You look down the street and see two people, Jones and Smith. Jones is walking about 20 paces ahead of Smith. They do not know each other. // Jones accidentally drops a $20 bill on the ground when he pulls his hand out of his pocket. He does not notice that he did this. // When Smith reaches the $20 bill, he picks it up and puts it in his pocket and then changes direction, clearly intending to keep the money.

You point out what you saw to the people at your table and get a few different responses.

For each person, please indicate how much their reaction to this event influences your impression of them. Assume that everyone is answering completely independently of one another - no one is giving a response based on what someone else said.

You ask someone at the table, Peters, whether he could do what Smith did. Peters says, "I would not keep the money. That said, I think I could choose to keep the money if I wanted to." How would this response influence your impression of Peters?

You ask someone at the table, Johnson, whether he could do what Smith did. Johnson says, "I would not keep the money. That said, I do not think I could choose to keep the money even if I wanted to." How would this response influence your impression of Johnson?

You ask someone at the table, Arnolds, if he could think that keeping the $20 is okay. Arnolds says, "I believe what Smith did was wrong. That said, if I wanted to, I think I could choose to believe that what Smith did was not wrong." How would this response influence your impression of Arnolds?

You ask someone at the table, Brown, if he could think that keeping the $20 is okay. Brown says, "I believe what Smith did was wrong. That said, if even if I wanted to, I do not think I could choose to believe that what Smith did was not wrong." How would this response influence your impression of Brown?

Appendix M

Statements used in the "beliefs in general" condition in Study 3.4

My current beliefs are ones that I voluntarily hold. Specifically, I could change what I believe if I wanted to even if this means I was being wrong or immoral by doing so.

My current job is one that I voluntarily hold. Specifically, I could change where I work if I wanted to even if this meant having a worse job.

My current habits are ones that I voluntarily engage in. Specifically, I could change my routines if I wanted to even if I had been doing them for years.

My current home is one that I voluntarily live in. Specifically, I could change where I live if I wanted to even if it meant changing many other parts of my life.

My behavior is voluntary. Specifically, when I make decisions to act a certain way, it is because I wanted to, and I could always act some other way if I wanted to even if this meant I was being immoral.

Appendix N

Statements used in the In General condition in Study 3.5

My current beliefs are ones that I voluntarily hold. Specifically, I could choose to hold different beliefs if I wanted to even if this meant being wrong or immoral.

My current job is one that I voluntarily hold. Specifically, I could change where I work if I wanted to even if this meant having a worse job.

My current home is one that I voluntarily live in. Specifically, I could change where I live if I wanted to even if it meant changing many other parts of my life.

My behavior is voluntary. Specifically, when I make decisions to act a certain way, it is because I want to, and I could act some other way even if this meant I was being immoral.

BIBLIOGRAPHY

Adams, R. M. (1985). Involuntary sins. *The Philosophical Review*, *94*(1), 3-31.

Alford, J. R., Hatemi, P. K., Hibbing, J. R., Martin, N. G., & Eaves, L. J. (2011). The politics of mate choice. *The Journal of Politics, 73*(2), 362-379.

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

Alston, W. P. (1988). The deontological conception of epistemic justification. *Philosophical Perspectives, 2*, 257-299.

Ames, D. R., & Johar, G. V. (2009). I'll know what you're like when I see how you feel: How and when affective displays influence behavior-based impressions. *Psychological Science*, *20*(5), 586-593. doi:10.1111/j.1467-9280.2009.02330.x

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: the effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*(2), 87-98. doi:10.1002/bdm.501

Band, E. B., & Weisz, J. R. (1988). How to feel better when it feels bad: Children's perspectives on coping with everyday stress. *Developmental Psychology*, *24*(2), 247-253. doi:10.1037/0012-1649.24.2.247

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bar-hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking & Reasoning*, *1*, 71-103. doi:10.1080/13546789508256906

Baron, J. (2008). *Thinking and deciding* (4 ed.). New York, NY: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml .2012.11.001

Bass, B., & Yammarino, F. (1991). Congruence of self and others' leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review*, 40, 437–454. doi:10.1111/j.1464-0597.1991.tb01002.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi: 10. 18637/jss.v067.i01.

Batson, C. D. (1990). How social an animal? The human capacity for caring. *American Psychologist*, *45*(3), 336-346. doi:10.1037/0003-066x.45.3.336

Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives. *Psychological Inquiry*, *2*, 107-122. doi:10.1207/s15327965pli0202_1

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin*, *20*(1), 3-19. doi:10.1177/0146167294201001

Bayles, M. (1982) Character, purpose and criminal responsibility. *Law and Philosophy, 1*, 5-20.

Beck, A. T., Brown, G. K., Steer, R. A., Kuyken, W., & Grisham, J. (2001). Psychometric properties of the beck self-esteem scales. *Behaviour Research and Therapy*, *39*(1), 115-124. doi:10.1016/s0005-7967(00)00028-0

Bierbrauer, G. (1979). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology*, *9*(1), 67-84. doi:10.1002/ejsp.2420090106

Blair, R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, *57*(1), 1-29. doi:10.1016/0010-0277(95)00676-p

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Brambilla, M., Hewstone, M., & Colucci, F. P. (2013). Enhancing moral virtues: Increased perceived outgroup morality as a mediator of intergroup contact effects. *Group Processes & Intergroup Relations*, *16*(5), 648-657. doi:10.1177/1368430212471737

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science, 23*(1), 27-34.
Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.

Brickman, P., Rabinowitz, V. C., Karuza Jr., J., Coates, D., Cohn, E., & Kidder, L. (1982). Models of helping and coping. *American Psychologist*, *37*(4), 368-384. doi:10.1037/0003-066x.37.4.368

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the “planning fallacy”: Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67, 366–381. doi:10.1037/0022-3514.67.3.366

Burger, J. M. (1986). Desire for control and the illusion of control: The effects of familiarity and sequence of outcomes. *Journal of Research in Personality*, *20*(1), 66-76. doi:10.1016/0092-6566(86)90110-8

Burger, J., & Cooper, H. (1979). The desirability of control. *Motivation and Emotion*, *3*(4), 381– 393. doi:10.1007/BF00994052

Chernyak, N., Kushnir, T., Sullivan, K. M., & Wang, Q. (2013). A comparison of American and Nepalese children's concepts of freedom of choice and social constraint. *Cognitive Science*, *37*(7), 1343-1355. doi:10.1111/cogs.12046

Cheung, B. Y., & Heine, S. J. (2015). The double-edged sword of genetic accounts of criminality: Causal attributions from genetic ascriptions affect legal decision making. *Pers Soc Psychol Bull*, *41*(12), 1723-1738.

Chignell, A. (2018). *The Ethics of Belief* (Spring 2017 ed. Vol. The Stanford Encyclopedia of Philosophy). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2018/entries/ethics-belief/

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Et, A. (1997). Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, *73*(3), 481-494. doi:10.1037//0022-3514.73.3.481

Cialdini, R. B., Schaller, M., Houlihan, D., Arps, K., Fultz, J., & Beaman, A. L. (1987). Empathy-based helping: Is it selflessly or selfishly motivated. *Journal of Personality and Social Psychology*, *52*(4), 749-758. doi:10.1037/0022-3514.52.4.749

Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F., & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of*

Personality and Social Psychology, *106*(4), 501-513. doi:10.1037/a0035880

Clifford, W. K. (1877). The ethics of belief. In T. Madigan (Ed.), *The ethics of belief and other essays* (pp. 70-96). Amherst, MA: Prometheus.

Coates, D. J., & Tognazzini, N. A. (2013). The contours of blame. *Blame: Its nature and norms*, 3-26.

Coates, D., Wortman, C.B., & Abbey, A. (1979) Reactions to victims, in I.H. Frieze , D. Bar-Tal & J.S. Carroll (eds) *New Approaches to Social Problems*. San Francisco, CA: Jossey-Bass

Cohen-Chen, S., Halperin, E., Saguy, T., & Zomeren, M. V. (2014). Beliefs about the malleability of immoral groups facilitate collective action. *Social Psychological and Personality Science*, *5*(2), 203-210. doi:10.1177/1948550613491292

Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology*, *81*(4), 697-710.

Cohen, N., & Ochsner, K. N. (2018). From surviving to thriving in the face of threats: The emerging science of emotion regulation training. *Current Opinion in Behavioral Sciences*, *24*, 143-155. doi:10.1016/j.cobeha.2018.08.007

Coyne, J. C., Kessler, R. C., Tal, M., Turnbull, J., Wortman, C. B., & Greden, J. F. (1987). Living with a depressed person. *Journal of Consulting and Clinical Psychology, 55*(3), 347-352. http://dx.doi.org/10.1037/0022-006X.55.3.347

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science, 4*, 308-315.

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology, 6,* 97-103.

Cushman, F. A. (2013). The role of learning in punishment, prosociality, and human uniqueness. In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (Eds.), *Cooperation and its Evolution* (pp. 333-372). Cambridge, Mass.: MIT Press.

D'Andrade, R. (1987). A folk model of the mind. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 112–148). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/ CBO9780511607660.006

Dar-Nimrod, I., Heine, S. J., Cheung, B. Y., & Schaller, M. (2011). Do scientific theories affect men's evaluations of sex crimes. *Aggressive Behavior*, *37*(5), 440-449. doi:10.1002/ab.20401

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, *47*(1), 1. doi:10.1086/268763

De Castella, K., Goldin, P., Jazaieri, H., Ziv, M., Dweck, C. S., & Gross, J. J. (2013). Beliefs about emotion: Links to emotion regulation, well-being, and psychological distress. *Basic and Applied Social Psychology*, *35*(6), 497-505. doi:10.1080/01973533.2013.840632

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Derakshan, N., Eysenck, M. W., & Myers, L. B. (2007). Emotional information processing in repressors: The vigilance–avoidance theory. *Cognition & Emotion*, *21*(8), 1585-1614. doi:10.1080/02699930701499857

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71-75. doi:10.1207/s15327752jpa4901_13

Dunkel-Schetter, C., & Skokan, L. A. (1990). Determinants of social support provision in personal relationships. *Journal of Social and Personal Relationships*, *7*(4), 437-450. doi:10.1177/0265407590074002

Dweck, C. S. (2013). *Self-theories*. Psychology Press. doi:10.4324/9781315783048

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, *30*(3), 133-140. doi:10.1257/jep.30.3.133

Epley, N., & Waytz, A. (2009). Mind perception. In S.T. Fiske, D.T. Gilbert, & G. Lindzey (Eds.), The handbook of social psychology (5th ed., pp. 498–541). New York: Wiley. doi:10.1002/9780470561119.socpsy001014

Fields, L., & Prinz, R. J. (1997). Coping and adjustment during childhood and adolescence. *Clinical Psychology Review*, *17*(8), 937-976. doi:10.1016/s0272-7358(97)00033-0

Fincham, F. D., & Jaspers, J. M. (1980) Attribution of responsibility: From man the scientist to man as lawyer. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13). New York: Academic Press

Finkel, E. J., Slotter, E. B., Luchies, L. B., Walton, G. M., & Gross, J. J. (2013). A brief intervention to promote conflict reappraisal preserves marital quality over time. *Psychological Science*, *24*(8), 1595-1601. doi:10.1177/0956797612474938

Ford, B. Q., & Gross, J. J. (2019). Why beliefs about emotion matter: An emotion-regulation perspective. *Current Directions in Psychological Science*, *28*(1), 74-81.

Ford, B. Q., & Troy, A. S. (2019). Reappraisal reconsidered: A closer look at the costs of an acclaimed emotion-regulation strategy. *Current Directions in Psychological Science*, *28*(2), 195-203. doi:10.1177/0963721419827526

Ford, B. Q., Lwi, S. J., Gentzler, A. L., Hankin, B., & Mauss, I. B. (2018). The cost of believing emotions are uncontrollable: Youths' beliefs about emotion predict emotion regulation and depressive symptoms. *Journal of Experimental Psychology: General*. doi:10.1037/xge0000396

Gal, D., & Rucker, D. D. (2010). When in doubt, shout!: Paradoxical influences of doubt on proselytizing. *Psychological Science*, *21*(11), 1701-1707. doi:10.1177/0956797610385953

Gebhardt, W. A., & Brosschot, J. F. (2002). Desirability of control: Psychometric properties and relationships with locus of control, personality, coping, and mental and somatic complaints in three Dutch samples. *European Journal of Personality, 16*, 423-438. doi:10.1002/per.463

Gibbs, J. L., Ellison, N. B., & Heino, R. D. (2006). Self-presentation in online personals: The role of anticipated future interaction, self-disclosure, and perceived success in internet dating. *Communication Research, 33*(2), 152-177.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107-119. doi:10.1037/0003-066x.46.2.107

Gilbert, D. T. (1993). The assent of man: Mental representation and the control of belief. In D. M. Wegner & J. W. Pennebaker (Eds.), *Century psychology series. Handbook of mental control* (pp. 57-87). Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological bulletin*, *117*(1), 21. doi:10.1037/0033-2909.117.1.21

Gilbert, D. T., Gill, M. J., & Wilson, T. D. (2002). The future is now: Temporal correction in affective forecasting. *Organizational Behavior and Human Decision Processes*, *88*(1), 430-444. doi:10.1006/obhd.2001.2982

Gill, M. J., & Cerce, S. C. (2017). He never willed to have the will he has: Historicist narratives, "civilized blame", and the need to distinguish two notions of free will. *Journal of Personality and Social Psychology, 112*(3), 361-382.

Gilovich, T. (1990). Differential construal and the false consensus effect. *Journal of Personality and Social Psychology*, *59*(4), 623-634. doi:10.1037/0022-3514.59.4.623

Gilovich, T., & Regan, D. T. (1986). The actor and the experiencer: Divergent patterns of causal attribution. *Social Cognition, 4,* 342–352. http://dx.doi.org/10.1521/soco.1986.4.3.342

Gino, F., Sharek, Z., & Moore, D. A. (2011). Keeping the illusion of control under control: Ceilings, floors, and imperfect calibration. *Organizational Behavior and Human Decision Processes*, *114*(2), 104-114. doi:10.1016/j.obhdp.2010.10.002

Golman, R., Loewenstein, G., Moene, K. O., & Zarri, L. (2016). The preference for belief consonance. *Journal of Economic Perspectives*, *30*(3), 165-188. doi:10.1257/jep.30.3.165

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In *Advances in Experimental Social Psychology: Advances in Experimental Social Psychology Volume 24* (pp. 319-359). doi:10.1016/s0065-2601(08)60333-0

Gromet, D. M., Goodwin, G. P., & Goodman, R. A. (2016). Pleasure from another's pain the influence of a target's hedonic states on attributions of immorality and evil.

Personality and Social Psychology Bulletin, 0146167216651408. doi:10.1177/0146167216651408

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271-299. doi:10.1037/1089-2680.2.3.271

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348-362. doi:10.1037/0022-3514.85.2.348

Guglielmo, S. & Malle, B. F. (2017). Information-acquisition processes in moral judgments of blame. *Personality and Social Psychology Bulletin, 43,* 957-971.

Haider-Markel, D. P., & Joslyn, M. R. (2008). Beliefs about the origins of homosexuality and support for gay rights: An empirical test of attribu- tion theory. *Public Opinion Quarterly, 72,* 291–310. http://dx.doi.org/10.1093/poq/nfn015

Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology, 33*(1), 1-36.

Halberstadt, A. G., Dunsmore, J. C., Bryant, A., Parker, A. E., Beale, K. S., & Thompson, J. A. (2013). Development and validation of the parents’ beliefs about children’s emotions questionnaire. *Psychological Assess*, *25*(4), 1195-1210. doi:10.1037/a0033695

Hammer, J. H., Cragun, R. T., Hwang, K., & Smith, J. M. (2012). Forms, frequency, and correlates of perceived anti-atheist discrimination. *Secularism and Nonreligion, 1*, 43-67.

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making*, *8*(3), 188-201.

Harris, P. L., Olthof, T., & Terwogt, M. M. (1981). Children's knowledge of emotion. *Journal of Child Psychology and Psychiatry*, *22*(3), 247-261. doi:10.1111/j.1469-7610.1981.tb00550.x

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252-264. doi:10.1207/s15327957pspr1003_4

Haslam, N., & Kvaale, E. P. (2015). Biogenetic explanations of mental disorder the mixed-blessings model. *Current Directions in Psychologi- cal Science, 24,* 399 – 404. http://dx.doi.org/10.1177/0963721415588082

Hauser, M., Cushman, F., Young, L., Kang-Xing, J.R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language*, *22*(1), 1-21. doi:10.1111/j.1468-0017.2006.00297.x

Heider, F. (1958). *The psychology of interpersonal relations*. Hoboken, NJ: John Wiley & Sons Inc.

Heider, F. (1958). *The psychology of interpersonal relations*. Hoboken, NJ, US: John Wiley & Sons Inc. doi:10.1037/10628-000

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Science*, *33*(2-3), 61-83; discussion 83. doi:10.1017/S0140525X0999152X

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin, 38*(1), 52-62.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, *76*(3), 349-366. doi:10.1037/0022-3514.76.3.349

Jaeger, J. (2017). r2glmm: Computes R squared for mixed (multilevel) models. R package version 0.1.2. https://CRAN.R-project.org/package=r2glmm

James, W. (1937). *The will to believe, and other essays in popular philosophy.* London: Longmans, Green and Co.

Johnson, J. T., Robinson, M. D., & Mitchell, E. B. (2004). Inferences about the authentic self: When do actions say more than mental states? *Journal of Personality and Social Psychology, 87,* 615–630. http://dx.doi.org/10 .1037/0022-3514.87.5.615

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1–24. doi:10.1016/0022-1031(67)90034-0

Jones, E. E., Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In Berkowitz, L. (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York, NY: Academic Press.

Kappes, A., & Schikowski, A. (2013). Implicit theories of emotion shape regulation of negative affect. *Cognition and Emotion*, *27*(5), 952-960. doi:10.1080/02699931.2012.753415

Karasawa, K. (1991). The effects of onset and offset responsibility on affects and helping judgments1. *Journal of Applied Social Psychology*, *21*(6), 482-499. doi:10.1111/j.1559-1816.1991.tb00532.x

Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.

Kay, A. C., Gaucher, D., McGregor, I., & Nash, K. (2010). Religious belief as compensatory control. *Personality and Social Psychology Review*, *14*(1), 37-48. doi:10.1177/1088868309353750

Kelley, H. H. (1971). *Attribution in social interaction*. Morristown, NJ: General Learning Press.

Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any "free" choice? Self and dissonance in two cultures. *Psychological Science*, *15*(8), 527-533. doi:10.1111/j.0956-7976.2004.00714.x

Klein, W. M., & Kunda, Z. (1992). Motivated person perception: Constructing justifications for desired beliefs. *Journal of experimental social psychology*, *28*(2), 145-168. doi:10.1016/0022-1031(92)90036-J

Kneeland, E. T., Nolen-Hoeksema, S., Dovidio, J. F., & Gruber, J. (2016a). Beliefs about emotion's malleability influence state emotion regulation. *Motivation and Emotion*, *40*(5), 740-749. doi:10.1007/s11031-016-9566-6

Kneeland, E. T., Nolen-Hoeksema, S., Dovidio, J. F., & Gruber, J. (2016b). Emotion malleability beliefs influence the spontaneous regulation of social anxiety. *Cognitive Therapy and Research*, *40*(4), 496-509. doi:10.1007/s10608-016-9765-1

Koehler, D. J., & Poon, C. S. K. (2006). Self-predictions overweight strength of current intentions. *Journal of Experimental Social Psychology*, 42, 517–524. doi: 10.1016/j.jesp.2005.08.003

Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition & Emotion*, *23*(1), 4-41. doi:10.1080/02699930802619031

Kouchaki, M., Smith, I. H., & Savani, K. (2018). Does deciding among morally relevant options feel like making a choice? How morality constrains people's sense of choice. *Journal of Personality and Social Psychology*. doi:10.1037/pspa0000128

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and" freezing.". *Psychological review*, *103*(2), 263. doi:10.1037/0033-295X.103.2.263

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, *108*(3), 480. doi:10.1037/0033-2909.108.3.480

Kvaale, E. P., Gottdiener, W. H., & Haslam, N. (2013). Biogenetic expla- nations and stigma: A meta-analytic review of associations among laypeople. *Social Science & Medicine, 96,* 95–103. http://dx.doi.org/10 .1016/j.socscimed.2013.07.017

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754-770.

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355-362. doi:10.1177/1948550617697177

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311-328. doi:10.1037/0022-3514.32.2.311

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist*, *46*(8), 819-834. doi:10.1037/0003-066x.46.8.819

Leary, M. R., Springer, C., Negel, L., Ansell, E., & Evans, K. (1998). The Causes, Phenomenology, and Consequences of Hurt Feelings. *Journal of Personality and Social Psychology*, *74*(5), 1225-1237.

Lebowitz, M. S., & Ahn, W.-K. (2014). Effects of biological explanations for mental disorders on clinicians' empathy. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 17786– 17790. http://dx.doi.org/10.1073/pnas.1414058111

Lilienfeld, S. O. (2017). Microaggressions: Strong claims, inadequate evidence. *Perspectives on Psychological Science*, *12*(1), 138-169. doi:10.1177/1745691616659391

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272-292. doi:10.1006/obhd.1996.0028

Lukinoff, G., & Haidt, J. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. New York, NY: Penguin Press.

MacDonald, T., & Ross, M. (1999). Assessing the accuracy of predictions about dating relationships: How and why do lovers' predictions differ from those made by observers? *Personality and Social Psychological Bulletin*, 25, 1417–1429. doi:10.1177/0146167299259007

Malle, B. F., & Knobe, J. (1997a). Which behaviors do people explain? A basic actor-observer asymmetry. *Journal of Personality and Social Psychology, 72*(2), 288-304.

Malle, B. F., & Knobe, J. (1997b). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*(2), 101-121.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*(2), 147-186.

Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: new answers to an old question. *Journal of Personality and Social Psychology*, *93*(4), 491-514. doi:10.1037/0022-3514.93.4.491

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, *102*(1), 72-90. doi:10.1037/0033-2909.102.1.72

Matute, H. (1996). Illusion of control: Detecting response-outcome independence in analytic but not in naturalistic conditions. *Psychological Science*, *7*(5), 289-293. doi:10.1111/j.1467-9280.1996.tb00376.x

Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, *26*(2-3), 131-146. doi:10.1080/01973533.2004.9646401

McArdle, M. (2018). Bias against conservatives works like any other prejudice. *Wall Street Journal*. Retrieved from https://www.washingtonpost.com/opinions/bias-against-conservatives-works-like-any-other-prejudice/2018/04/10/17fa1838-3c40-11e8-974f-aacd97698cef_story.html

McFarland, C., & Miller, D. T. (1990). Judgments of self-other similarity. *Personality and Social Psychology Bulletin*, *16*(3), 475-484. doi:10.1177/0146167290163006

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences*, *20*(1), 91-102; discussion 103. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10096996

Meyer, J. P., & Mulherin, A. (1980). From attribution to helping: An analysis of the mediating effects of affect and expectancy. *Journal of Personality and Social Psychology*, *39*(2), 201-210. doi:10.1037/0022-3514.39.2.201

Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology*, *53*(2), 298-305. doi:10.1037/0022-3514.53.2.298

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, *74*(1), 53-62. doi:10.1037/0022-3514.74.1.53

Miller, J. G., Das, R., & Chakravarthy, S. (2011). Culture and the role of choice in agency. *Journal of Personality and Social Psychology*, *101*(1), 46-61. doi:10.1037/a0023330

Mitmansgruber, H., Beck, T. N., Höfer, S., & Schüßler, G. (2009). When you don't like what you feel: Experiential avoidance, mindfulness and meta-emotion in emotion regulation. *Personality and Individual Differences, 46,* 448–453. http://dx.doi.org/10.1016/j.paid.2008.11.013

Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General, 146,* 123–133. http://dx.doi.org/ 10.1037/xge0000234

Monroe, A. E., & Reeder, G. D. (2011). Motive-matching: Perceptions of intentionality for coerced action. *Journal of Experimental Social Psychology, 47*, 1255-1261.

Monterosso, J., Royzman, E. B., & Schwartz, B. (2005). Explaining away responsibility: Effects of scientific explanation on perceived culpability. *Ethics & Behavior, 15*(2), 139-158.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*(6), 949-971. doi:10.1037/0022-3514.67.6.949

Murray, S. L., Holmes, J. G., Bellavia, G., Griffin, D. W., & Dolderman, D. (2002). Kindred spirits? The benefits of egocentrism in close relationships. *Journal of Personality and Social Psychology, 82*(4), 563-581.

Nadal, K. L. (2011). The Racial and Ethnic Microaggressions Scale (REMS): Construction, reliability, and validity. *Journal of Counseling Psychology, 58*, 470-480. doi:10.1037/a0025193

Nadler, J., & McDonnell, M.-H. (2012). Moral character, motive, and the psychology of blame. *Cornell Law Review, 97*, 255–304.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133‑142.

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation, Vol 26* (pp. 125-173). Elsevier. doi:10.1016/s0079-7421(08)60053-5

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259. doi:10.1037/0033-295x.84.3.231

Niven, K., Totterdell, P., & Holman, D. (2009). A classification of controlled interpersonal affect regulation strategies. *Emotion*, *9*(4), 498-509. doi:10.1037/a0015962

Nolan, H. (2018). Ideology is a choice. *Splinter*. Retrieved from https://splinternews.com/ideology-is-a-choice-1825172619

Ochsner, K. N., & Gross, J. J. (2008). Cognitive emotion regulation. *Current Directions in Psychological Science*, *17*(2), 153-158. doi:10.1111/j.1467-8721.2008.00566.x

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fmri study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, *14*(8), 1215-1229. doi:10.1162/089892902760807212

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141-162. doi:10.1016/j.cognition.2015.06.010

Pancer, S. M., Adams, D. A., Mollard, D., Solsberg, D., & Tammen, L. (1979). Perceived distinctiveness of the handicapped. *The Journal of Social Psychology*, *108*(2), 275-276. doi:10.1080/00224545.1979.9711645

Paolacci, G., & Chandler, J. (2014). Inside the turk. *Current Directions in Psychological Science*, *23*(3), 184-188. doi:10.1177/0963721414531598

Pascal, B. (1852). *Pensées*. Dezobry et E. Magdeleine.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In *Communication and Persuasion* (pp. 1-24). New York, NY: Springer New York. doi:10.1007/978-1-4612-4964-1_1

Pew Research Center. (2015). Teen, social media and technology.

Pew Research Center. (2015). U.S. public becoming less religious.

Pew Research Center. (2016). The new food fights: U.S. public divides over food science.

Pew Research Center. (2016). The politics of climate.

Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In P. Shaver & M. Mikulincer (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91– 108). Washington, DC: American Psychological Association.

Pizarro, D. A., Tannenbaum, D., & Uhlmann, E. (2012). Mindless, harmless, and blameworthy. *Psychological Inquiry*, *23*(2), 185-188.

Pronin, E. (2009). The introspection illusion. In *Advances in Experimental Social Psychology Volume 8* (pp. 1-67). doi:10.1016/S0065-2601(08)00401-2

Pronin, E., & Kugler, M. B. (2010). People believe they have more free will than others. *Proceedings of the National Academy of Sciences*, *107*(52), 22469-22474. doi:10.1073/pnas.1012046108

Pronin, E., Fleming, J. J., & Steffel, M. (2008). Value revelations: Disclosure is in the eye of the beholder. *Journal of Personality and Social Psychology*, *95*(4), 795-809. doi:10.1037/a0012710

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological review*, *111*(3), 781-799. doi:10.1037/0033-295X.111.3.781

Pronin, E., Kruger, J., Savtisky, K., & Ross, L. (2001). You don’t know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology*, *81*(4), 639. doi:10.1037/0022-3514.81.4.639

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381. doi:10.1177/0146167202286008

Reeder, G. D. (1993). Trait-behavior relations and dispositional inference. *Personality and Social Psychology Bulletin*, *19*(5), 586-593.

Reeder, G. D. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry, 20,* 1–18. http://dx.doi.org/10.1080/10478400802615744

Reeder, G. D., Monroe, A. E., & Pryor, J. B. (2008). Impressions of Milgram's obedient teachers: Situational cues inform inferences about motives and traits. *Journal of Personality and Social Psychology*, *95*(1), 1-17. doi:10.1037/0022-3514.95.1.1

Risucci, D. A., Tortolani, A. J., & Ward, R. J. (1989). Ratings of surgical residents by self, supervisors and peers. *Surgical Gynecology and Obstetrics*, 169, 519–526.

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology, 68*(3), 404-417. doi:10.1037/0022-3514.68.3.404

Rogers, T., Moore, D. A., & Norton, M. I. (2017). The belief in a favorable future. *Psychological Science, 28*(9), 1290–1301. doi:10.1177/0956797617706706

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *The*

Jean Piaget symposium series. Values and knowledge (pp. 103-135). US: Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sabini, J., & Silver, M. (1998). *Emotion, character, and responsibility*. New York, NY: Oxford University Press.

Sankowski, E. (1977). Responsibility of persons for their emotions. *Canadian Journal of Philosophy, 7*(4), 829-840.

Savani, K., Markus, H. R., Naidu, N. V., Kumar, S., & Berlia, N. (2010). What counts as a choice? U.S. Americans are more likely than Indians to construe actions as choices. *Psychol Sci*, *21*(3), 391-398. doi:10.1177/0956797609359908

Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a bayesian model of theory of mind. *Current opinion in psychology*, *17*, 15-21. doi:10.1016/j.copsyc.2017.04.019

Schlesinger, I. M. (1992). The experiencer as an agent. *Journal of Memory and Language, 31,* 315–332. http://dx.doi.org/10.1016/0749-596X (92)90016-Q

Schroder, H. S., Dawood, S., Yalch, M. M., Donnellan, M. B., & Moser, J. S. (2015). The role of implicit theories in mental health symptoms, emotion regulation, and hypothetical treatment choices in college stu- dents. *Cognitive Therapy and Research, 39,* 120–139. http://dx.doi.org/ 10.1007/s10608-014-9652-6

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology*, *84*(1), 60-79. doi:10.1037/0022-3514.84.1.60

Seligman, M. E. P. (1974). Depression and learned helplessness. In R. J. Friedman & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 83-113). Washington, DC: Winston-Wiley.

Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco, CA: Freeman.

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

Shaw, L. L., Batson, C. D., & Todd, R. M. (1994). Empathy avoidance: Forestalling feeling for another in order to escape the motivational consequences. *Journal of Personality and Social Psychology*, *67*(5), 879-887. doi:10.1037/0022-3514.67.5.879

Sheppes, G., & Meiran, N. (2008). Divergent cognitive costs for online forms of reappraisal and distraction. *Emotion*, *8*(6), 870-874. doi:10.1037/a0013711

Sheppes, G., Scheibe, S., Suri, G., & Gross, J. J. (2011). Emotion-regulation choice. *Psychol Sci*, *22*(11), 1391-1396. doi:10.1177/0956797611418350

Sheppes, G., Scheibe, S., Suri, G., Radu, P., Blechert, J., & Gross, J. J. (2014). Emotion regulation choice: A conceptual framework and supporting evidence. *J Exp Psychol Gen*, *143*(1), 163-181. doi:10.1037/a0030831

Shultz, T. R., Schleifer, M., & Altman, I. (1981). Judgments of causation, responsibility, and punishment in cases of harm-doing. *Canadian Journal of Behavioural Science, 13*(3), 238-253.

Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology*, *68*(2), 247-260. doi:10.1037/0022-3514.68.2.247

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2014). Anchoring is not a false-positive: Maniadis, Tufano, and List's (2014) 'Failure-to-Replicate' is actually entirely consistent with the original. *SSRN Electronic Journal*. doi:10.2139/ssrn.2351926

Skerry, A. E., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Curr Biol*, *25*(15), 1945-1954. doi:10.1016/j.cub.2015.06.009

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895-917.

Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, *115*(2), 268-281. doi:10.1016/j.cognition.2009.12.017

Smith, A. M. (2008). Control, responsibility, and moral assessment. *Philosophical Studies, 138*(3), 367-392.

Smith, H. M. (2011). Non-tracing cases of culpable ignorance. *Criminal Law and Philosophy, 5*(2), 115-146.

Spector, P. E., Sanchez, J. I., Siu, O. L., Salgado, J., & Ma, J. (2004). Eastern versus western control beliefs at work: An investigation of secondary control, socioinstrumental control, and work locus of control in china and the US. *Applied Psychology*, *53*(1), 38-60. doi:10.1111/j.1464-0597.2004.00160.x

Sprecher, S., & Hendrick, S. S. (2004). Self-disclosure in intimate relationships: Associations with individual and relationship characteristics over time. *Journal of Social and Clinical Psychology, 23*(6), 857-877.

Ståhl, T., & van Prooijen, J.W. (2018). Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. *Personality and Individual Differences*, *122*, 155-163. doi:10.1016/j.paid.2017.10.026

Ståhl, T., Zaal, M. P., & Skitka, L. J. (2016). Moralized rationality: Relying on logic and evidence in the formation and evaluation of belief can be seen as a moral issue. *PLOS ONE*, *11*, e0166332. doi:10.1371/journal.pone.0166332

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342-357. doi:10.1037/0022-0663.89.2.342

Steup, M. (2017). Believing intentionally. *Synthese*, *194*(8), 2673-2694. doi:10.1007/s11229-015-0780-7

Suhay, E., Brandt, M. J., & Proulx, T. (2017). Lay belief in biopolitics and political prejudice. *Social Psychological and Personality Science, 8,* 173–182. http://dx.doi.org/10.1177/1948550616667615

Suri, G., Sheppes, G., Young, G., Abraham, D., Mcrae, K., & Gross, J. J. (2018). Emotion regulation choice: The role of environmental affordances. *Cogn Emot*, *32*(5), 963-971. doi:10.1080/02699931.2017.1371003

Szczurek, L., Monin, B., & Gross, J. J. (2012). The stranger effect: The rejection of affective deviants. *Psychological Science*, *23*(10), 1105-1111. doi:10.1177/0956797612445314

Tamir, M. (2009). Differential preferences for happiness: Extraversion and trait-consistent emotion regulation. *Journal of Personality*, *77*(2), 447-470. doi:10.1111/j.1467-6494.2008.00554.x

Tamir, M., Bigman, Y. E., Rhodes, E., Salerno, J., & Schreier, J. (2015). An expectancy-value model of emotion regulation: Implications for motivation, emotional experience, and decision making. *Emotion*, *15*(1), 90-103. doi:10.1037/emo0000021

Tamir, M., John, O. P., Srivastava, S., & Gross, J. J. (2007). Implicit theories of emotion: Affective and social outcomes across a major life transition. *Journal of Personality and Social Psychology*, *92*(4), 731-744. doi:10.1037/0022-3514.92.4.731

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193-210. doi:10.1037/0033-2909.103.2.193

Taylor, S. E., Kemeny, M. E., Reed, G. M., Bower, J. E., & Gruenewald, T. L. (2000). Psychological resources, positive illusions, and health. *American Psychologist*, *55*(1), 99-109. doi:10.1037/0003-066x.55.1.99

Tetlock, P. E., Kristel, O. V., Beth, S., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853-870. doi:10.1037/0022-3514.78.5.853

Thomas, K. R. (2008). Macrononsense in multiculturalism. *Am Psychol*, *63*(4), 274-5; discussion 277. doi:10.1037/0003-066X.63.4.274

Torres, L., Driscoll, M. W., & Burrow, A. L. (2010). Racial microaggressions and psychological functioning among highly achieving african-americans: A mixed-methods approach. *Journal of Social and Clinical Psychology*, *29*(10), 1074-1099. doi:10.1521/jscp.2010.29.10.1074

Troy, A. S., Ford, B. Q., Mcrae, K., Zarolia, P., & Mauss, I. B. (2017). Change the things you can: Emotion regulation is more beneficial for people from lower than from higher socioeconomic status. *Emotion*, *17*(1), 141-154. doi:10.1037/emo0000210

Troy, A. S., Shallcross, A. J., & Mauss, I. B. (2013). A person-by-situation approach to emotion regulation: Cognitive reappraisal can either help or hurt, depending on the context. *Psychol Sci*, *24*(12), 2505-2514. doi:10.1177/0956797613496434

Tullett, A. M., & Plaks, J. E. (2016). Testing the link between empathy and lay theories of happiness. *Personality and Social Psychology Bulletin*, *42*(11), 1505-1521. doi:10.1177/0146167216665092

Turri, J., Rose, D., & Buckwalter, W. (2017). Choosing and refusing: Doxastic voluntarism and folk psychology. *Philosophical Studies*, 1-31.

Turri, J., Rose, D., & Buckwalter, W. (2018). Choosing and refusing: Doxastic voluntarism and folk psychology. *Philosophical Studies*, *175*, 2507-2537. doi:10.1007/s11098-017-0970-x

Uhlmann, E. L., & Zhu, L. [. L. (2013). Acts, persons, and intuitions: Person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science*, *x*, xxx-xxx. doi:10.1177/1948550613497238

Uhlmann, E. L., Zhu, L. L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, *126*(2), 326-334.

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, *29*(9), 1159-1168. doi:10.1177/0146167203254597

Van Dillen, L. F., & Koole, S. L. (2007). Clearing the mind: A working memory model of distraction from negative mood. *Emotion*, *7*(4), 715-723. doi:10.1037/1528-3542.7.4.715

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049-1062. doi:10.1037/0022-3514.67.6.1049

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*(1), 34-52. doi:10.1037/0033-295x.101.1.34

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation. Sources of the experience of will. *American Psychologist*, *54*(7), 480-492. doi:10.1037/0003-066X.54.7.480

Weiner, B. (1985). Attribution Theory. In *Human Motivation* (pp. 275-326). New York, NY: Springer New York. doi:10.1007/978-1-4612-5092-0_7

Weiner, B. (1993). On sin versus sickness: A theory of perceived responsibility and social motivation. *American Psychologist*, *48*(9), 957-965. doi:10.1037/0003-066x.48.9.957

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

Weiner, B., Graham, S., & Chandler, C. (1982). Pity, anger, and guilt. *Personality and Social Psychology Bulletin*, *8*(2), 226-232. doi:10.1177/0146167282082007

Weiner, B., Perry, R. P., & Magnusson, J. (1988). An attributional analysis of reactions to stigmas. *Journal of Personality and Social Psychology*, *55*(5), 738-748. doi:10.1037/0022-3514.55.5.738

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*(1), 59-91. doi:10.1146/annurev.psych.51.1.59

West, K. (2019). Testing hypersensitive responses: Ethnic minorities are not more sensitive to microaggressions, they just experience them more frequently. *Pers Soc Psychol Bull*, 146167219838790. doi:10.1177/0146167219838790

White, R. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*, 297–330. doi:10.1037/h0040934

Williams, W. C., Morelli, S. A., Ong, D. C., & Zaki, J. (2018). Interpersonal emotion regulation: Implications for affiliation, perceived support, relationships, and well-being. *Journal of Personality and Social Psychology*, *115*(2), 224-254. doi:10.1037/pspi0000132

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117-142. doi:10.1037/0033-2909.116.1.117

Wood, T., & Porter, E. (2018). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*. doi:10.1007/s11109-018-9443-y

Wortman, C. B., & Brehm, J. W. (1975). Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In L. Berkowitz (Ed.) *Advances in Experimental Social Psychology Volume 8* (pp. 277-336). San Diego, CA: Academic Press. doi:10.1016/s0065-2601(08)60253-1

Wortman, C. B., & Lehman, D. R. (1985). Reactions to victims of life crises: Support attempts that fail. In I.G. Sarason & B.R. Sarason (eds) *Social Support: Theory, Research and Applications* (pp. 463-489). Dordrecht, Netherlands: Martinus Nijhoff. doi:10.1007/978-94-009-5115-0_24

Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology*, *115*(6), 929-943.