

# Lay Beliefs About the Controllability of Everyday Mental States

Corey Cusimano and Geoffrey P. Goodwin  
University of Pennsylvania

Prominent accounts of folk theory of mind posit that people judge others' mental states to be uncontrollable, unintentional, or otherwise involuntary. Yet, this claim has little empirical support: few studies have investigated lay judgments about mental state control, and those that have done so yield conflicting conclusions. We address this shortcoming across six studies, which show that, in fact, lay people attribute to others a high degree of intentional control over their mental states, including their emotions, desires, beliefs, and evaluative attitudes. For prototypical mental states, people's judgments of control systematically varied by mental state category (e.g., emotions were seen as less controllable than desires, which in turn were seen as less controllable than beliefs and evaluative attitudes). However, these differences were attenuated, sometimes completely, when the content of and context for each mental state were tightly controlled. Finally, judgments of control over mental states correlated positively with judgments of responsibility and blame for them, and to a lesser extent, with judgments that the mental state reveals the agent's character. These findings replicated across multiple populations and methods, and generalized to people's real-world experiences. The present results challenge the view that people judge others' mental states as passive, involuntary, or unintentional, and suggest that mental state control judgments play a key role in other important areas of social judgment and decision making.

*Keywords:* blame, control, intentionality, mental states, responsibility

*Supplemental materials:* <http://dx.doi.org/10.1037/xge0000547.supp>

Consider someone who asks for a raise because she wants to buy a new house. Observers would likely judge this act as under her control, in that she intentionally chose to make the request and could have chosen not to make it if she had so desired. But, would they also judge that she had control over whether she *wanted* to buy a new house, or whether she *believed* that asking for a raise was the way to accomplish this goal? According to many scholars, the answer would likely be “no,” reflecting a widely held thesis that ordinary people judge mental states as unintentional, involuntary occurrences that are akin to things that simply happen to people rather than things that they do. However, as we will review, the existing evidence for this viewpoint is indirect and incomplete, and therefore insufficient to answer the central questions of this

paper: How much control do ordinary people generally attribute to others over their mental states? Do they in fact judge others to have little to no control over their mental states? And do attributions of mental state control depend on which aspect of mental functioning people are judging? Our goal in this paper is to answer these questions using several comprehensive and triangulating methodologies.

## The Importance of Perceived Intentionality and Control

Attributions of behavioral control have been widely studied within social psychology, and have been shown to affect an array of judgments and decisions (Dweck & Leggett, 1988; Heider, 1958; Skinner, 1996; Weiner, 1995). For instance, when people judge that others have control over their behavior they are more inclined to hold them responsible for it (Alicke, 2000; Weiner, 1995), and to try to modify it (Dweck, Chiu, & Hong, 1995; Dweck & Leggett, 1988). Judgments of intentionality, a more specific construct of control (Malle & Knobe, 1997a), play an especially critical role in shaping explanations of behavior and judgments of responsibility (Heider, 1958; Malle, 2004; Malle, Guglielmo, & Monroe, 2014). Intentional norm violations, in contrast to unintentional norm violations, prompt observers to search for reasons underlying the behavior (Malle et al., 2014; Monroe & Malle, 2017), are deemed more blameworthy (see Malle et al., 2014), and are more likely to induce inferences of negative character (Reeder, 2009).

Although not as widely studied, there is also some existing evidence that people incorporate controllability and intentionality information when evaluating mental states. For instance, a seminal study

---

Corey Cusimano and Geoffrey P. Goodwin, Department of Psychology, University of Pennsylvania.

Portions of this work have been previously presented at the International Conference of Thinking (2016). Studies 1, 3, and 4a were partially reported in the published proceedings for the 39th Annual Conference for the Cognitive Science Society (Cusimano & Goodwin, 2017b). Data and analysis scripts can be accessed at <https://osf.io/hcsnf/>.

The authors thank Jon Baron, Bertram Malle, and Dan Swingley for feedback on this research. The authors also thank Luiza Santos for serving as second coder for Study 2. Funding for this research was provided by the Norman Anderson Fellowship to Corey Cusimano.

Correspondence concerning this article should be addressed to Corey Cusimano, Department of Psychology, University of Pennsylvania, Solomon Labs, 3720 Walnut Street, Philadelphia, PA 19104. E-mail: [cusimano@sas.upenn.edu](mailto:cusimano@sas.upenn.edu)

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). And more broadly, it has been shown that people who view anxiety, schizophrenia, and other mental illnesses as controllable are more likely to blame someone for having one of these disorders (Lebowitz & Ahn, 2014).

People's attributions of control also seem to predict their motivations, strategies, and sense of responsibility with respect to their own emotions. 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).

There is thus a considerable body of literature indicating the importance of control for a variety of social processes. For behaviors, control plays an especially important role in guiding judgments of responsibility and blame, and in shaping attempts at modification. And, although the evidence is less complete for mental states, control seems to occupy a similar role for them too. However, despite evidence about the importance of control for social judgment, we know surprisingly little about the degree of control people attribute over mental states in the first place. In fact, there is no robust evidence indicating whether commonplace desires, beliefs, emotions, and so forth, are seen as generally controllable or uncontrollable.

This is an important gap because it is clear that mental states are frequently thought about, evaluated, and reacted to in their own right. That is, in addition to caring about others' *conduct*, people also care inherently about what others *think*, *feel*, and *want*. People wonder about why others have the attitudes that they do (Malle & Knobe, 1997b), want others to think positively of them (Leary, Springer, Negel, Ansell, & Evans, 1998), want those around them to share their beliefs and values (Haidt, Rosenberg, & Hom, 2003; Pronin, Fleming, & Steffel, 2008; Tetlock, Kristel, Elson, Green, & Lerner, 2000), and want those in their community to feel culturally appropriate emotions (e.g., Cohen & Rankin, 2004; Eid & Diener, 2001; Sankowski, 1977). When others do have discrepant attitudes, people are typically motivated to try to change them (Akhtar & Wheeler, 2016). In sum, mental states are often themselves the objects of social judgment, which provides impetus for better understanding how controllable they are generally judged to be. Accordingly, the major goal of the present paper is to provide this understanding across a wide range of everyday mental states.

### Prior Work on Perceived Mental State Control

The dominant assumption among psychologists appears to be that mental states are perceived as passively experienced rather than actively intended or controlled. 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 (1997b) 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). Subsequent work in behavior explanation has similarly assumed that many mental states are unintentional (e.g., they are explained in the same way that unintentional behaviors are explained, Malle, 2004, 2014).

The assumption of low perceived mental state control is also prominent in philosophical accounts of everyday mental state evaluation (see, e.g., Adams, 1985; Arpaly & Schroeder, 2013; Sabini & Silver, 1998; A. Smith, 2005, 2008; H. Smith, 2011). For instance, in his seminal paper, Adams (1985) counts among so-called "involuntary sins," "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" (p. 4, emphasis added). But as we review below, the thesis that people view mental states as uncontrollable, unintentional, or involuntary is not empirically supported. Indeed, no straightforward conclusion can be drawn from the existing evidence due to its piecemeal and highly conflicting nature.

Some of the earliest work on folk conceptions of mental state controllability emerged from investigations of the relationship between grammar and the semantics of agency, causation and force. Several theorists argued that mental verbs (e.g., "fear" or "like") denote involuntary states, as opposed to voluntary actions (e.g., "kill"), and that this fact was revealed through the ways people naturally talk about them (see, e.g., Miller & Johnson-Laird, 1976). For instance, in an early article, Katz and Postal (1964) argued that the issuing of a command is constrained by whether the recipient of the command can willfully choose to carry it out. Because mental state verbs are often infelicitous when expressed as commands (e.g., "Want to go!" or "Please attract me"), Katz and Postal concluded that "being in such psychological states as belief, understanding, wanting and hoping is not subject to a person's will" (p. 77). This conclusion was highly influential in linguistics and psycholinguistics (see, e.g., Brown & Fish, 1983, and Rudolph & Försterling, 1997); however, the data are not as clear as these scholars supposed. Although some mental states resist the imperative form, there are plenty of instances where they are perfectly grammatical (e.g., "Be happy!", "Just believe I'm telling you the truth!", or "Hope for the best!"). This observation led others to conclude that, in fact, mental states were "subject to some degree of voluntary control" (Huddleston, 1970, p. 508). Thus, even if grammaticality judgments provide an appropriate means of assessing perceptions of control in everyday settings, the evidence they provide is conflicting.

More recent work has eschewed these indirect methods of assessing mental state control by instead asking people directly to judge how controllable or intentional mental states are. Unfortunately, these more direct methods have not yielded clearer conclusions.

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 position. 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 (where perceived control would not be inflated by syntactic cues). 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. Relatedly, because the questions were closely matched to the contents of the vignettes, it is possible that subjects reported agreement with the follow-up statements because they appeared almost like comprehension checks (testing whether they read the vignette closely) rather than assessments of their true beliefs. A critic may therefore wonder whether the opposite result would have held if the vignette stated initially that the person could *not* choose to change their belief (with the follow-up question matching that same negative content). 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), suggest-

ing 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 their experiences.<sup>1</sup> 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 (1997a) 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.

### Control as It Pertains to Different Mental State Kinds

An additional shortcoming of past empirical research on this topic is that scholars have theorized about control over mental states as if “mental states” comprise a single category, or have otherwise developed coarse binary classifications such as “experiences” and “thoughts.” However, people’s lay model of the mind

<sup>1</sup> Arguably, the choice ratings actually undermine Gilovich and Regan’s conclusion regarding mental states, because the mean rating for experiences was 4.46 on a 7-point scale, thereby indicating more choice than not (actions were rated at 5.54).

contains a variety of mental state concepts distinguished by their role in cognition as well as their predictive and explanatory functions (D'Andrade, 1987; Heider, 1958; Malle, 2004; Wellman, 1990). It is therefore possible that perceived control might vary according to the type of mental state being evaluated (D'Andrade, 1987). Intrigued by this possibility, we sought to gain a better understanding of people's lay beliefs about mental state control by studying how much control people attribute across several distinct categories of mental states. We focused on those states that have historically played important roles in social cognition: *emotions*, *beliefs*, *desires*, and *evaluations*. We briefly describe each of these categories below, alongside prominent theoretical claims about their controllability.

### Emotions

Emotions are feeling states with a subjective valence that are typically accompanied by characteristic somatic events, including facial expressions, as well as characteristic action tendencies (Oatley & Johnson-Laird, 1987; Johnson-Laird & Oatley, 1989; Russell, 1983). Emotions (and evaluations, see below) are treated as the prototypical example of unintentional and uncontrollable experiences (Adams, 1985; Malle & Knobe, 1997b, quoted above). This view is succinctly articulated by Sabini and Silver (1998) who write, "when people describe an experience as emotional they imply that it is to an extent beyond the will, passive, a reaction as opposed to an action" (p. 5). But it is difficult to reconcile this view with recent findings from the emotion regulation literature suggesting that people attribute some degree of control to *themselves* over what they feel. Even in the emotion regulation literature, however, scholars have focused on variation in perceived control between individuals, rather than on prototypical attributions of control over emotions, or on attributions of control relative to other mental states.

### Desires

A desire is a wanting or wishing for something to happen or for some state of affairs to be different. Desires are commonly denoted by words such as "want," "crave," "wish for," and "hope." As in the case of emotions, scholars have sometimes attributed to lay individuals the belief that desires are uncontrollable, expressed as the belief that, "there seems to be no way to make oneself not want something or to want something one has no desire for" (Adams, 1985; D'Andrade, 1987, p. 119). According to some scholars, desires are judged to be the least controllable of mental states (i.e., even less controllable than emotions, D'Andrade, 1987, p. 117).

### Beliefs

A belief is an assent to or acceptance of some proposition as true. Although many scholars argue that beliefs are, as a matter of fact, uncontrollable (e.g., Alston, 1988; Epley & Gilovich, 2016; James, 1937), many also assume that ordinary people view beliefs as controllable (D'Andrade, 1987). Turri et al. (2018) provide evidence that ordinary people view beliefs as voluntarily chosen, though the sampling of beliefs complicate the interpretation of this evidence (see above). Similar to D'Andrade (1987), Turri et al. (2018) suggest that beliefs are judged to be among the most controllable mental states.

### Evaluations

Evaluations, also sometimes referred to as "valuings" (e.g., Malle, 2004) or "attitudes" (e.g., Hartshorne et al., 2016), reflect some positive or negative evaluation of someone or something (e.g., "loving," "disliking," "valuing," "caring," "trusting," etc.). Evaluations share properties with beliefs, desires, and emotions, but can be distinguished from each of them. They tend to be judged as more habitual or long-lasting than emotions (Hartshorne et al., 2016), and also as more abstract (e.g., unlike emotions, evaluative attitudes are rarely described as being instantiated in a concrete time and place; Pylikkanen, 1999). Unlike desires, they do not pertain to an unfulfilled state. And unlike beliefs, they do not represent any particular proposition as true or false (see Malle, 2004, for a discussion). Scholars typically do not distinguish evaluations from emotions in discussions of mental state control. For instance, D'Andrade (1987), while discussing folk concepts of "feelings and emotions," conflates evaluations with emotions when he writes, "according to the folk model, one cannot will one's self to hate or not to hate, to love or not to love someone, or even to enjoy something" (p. 119). Thus, claims about the perceived controllability of evaluations often dovetail with claims about the controllability of emotions; namely, that they (evaluations and emotions) are treated as unintentional or involuntary occurrences (see also Adams, 1985; A. Smith, 2005; H. Smith, 2011).

This categorization of mental states is not exhaustive, but it captures the primary mental state classes that have been referred to in accounts of theory of mind. Other mental states that people often attribute to others include perceptions, memories, attention, and intentions. There are also important subdivisions within each of our main classes (e.g., anger vs. sadness, impression vs. belief). We will discuss the relevance of these other divisions to lay beliefs about mental control in the General Discussion.

### Overview of Research

To review, despite the apparent importance of mental state control, and the preponderance of claims stating that people view mental states as unintentional and uncontrollable occurrences, there has not yet been a broad, systematic empirical investigation of whether people view the mind and its various components as generally controllable or not. The studies that have been conducted were not originally designed to answer this basic question. Accordingly, they suffer from various limitations, and have come to conflicting conclusions.

In light of these gaps and inconsistencies, we had two primary goals in the studies that follow. First, we examined whether mental states belonging to each of the classes listed above are typically seen as uncontrollable or controllable, and unintentional or intentional. Given that people tend to rate their own emotions as somewhat controllable, and that emotions are widely considered to be among the least controllable mental states, we expected (contrary to many claims cited above) that people would attribute considerable control to others over their mental states. Our second goal was to test for systematic differences *between* mental state categories regarding their perceived controllability, intentionality, and changeability (e.g., emotions relative to beliefs, desires, etc.). Consistent with prior theorizing, we predicted that emotions would be judged as less controllable than other mental states, but we did

not make any predictions regarding the relative controllability and intentionality of desires, beliefs, and evaluations.

Our studies are grouped into two sets. The first set of studies (1–3) were exploratory investigations of the degree of control people attribute to others over various mental states. Using three distinct methods, these studies yielded a reliable ordering of control across mental state types. Studies 4–6 were more stringent tests of whether this ordering held up when the content and surrounding context of each mental state were tightly controlled.

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.

To increase the likelihood that our findings apply to everyday mental state evaluation, we used subject-generated stimuli (Studies 1 and 2), as well as scenarios that specified the surrounding context clearly (Studies 2, 4, 5, 6). These features represent important methodological improvements over the studies cited above, which relied heavily on unrepresentative, sparsely described, and hypothetical scenarios; although we do rely on abstract (Study 3) and hypothetical scenarios (Studies 4–6) to some extent as well. Finally, because there is reason to think that control judgments inform other important social judgments (e.g., Weiner, 1995; see discussion above), we also conducted preliminary tests of whether variation in perceived control for everyday mental states correlates with judgments of agents' responsibility (Studies 1, 3, 6) and blame for them (Studies 2 and 5), as well as with judgments that the mental state reveals the agent's character (Studies 1, 2, 5, 6).

### Transparent Reporting

For all studies we report all measures and conditions. We preregistered Studies 4 and 5. Sample sizes for all studies were set prior to data collection. Materials, data, analysis code, and links to preregistration documents are available on the Open Science Framework (<https://osf.io/hcsnf/>). We did not exclude any subjects from our analyses. All studies reported in this paper were approved by the Institutional Review Board at the University of Pennsylvania.

### Study 1

Study 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. Our primary goal was to assess and compare ratings of control across eight main categories of mental states.

### 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 responses. Any response that more than one subject provided was automatically included. For items that did not produce five pairs of duplicate responses, we selected nonduplicate responses by attempting to maximize the differences in content among the set of responses. 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 to 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 shows means and standard deviations for each category (mental states and observable behavior foils).

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. 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  $p_a = 1$ . The algorithm and R code for calculating adjusted  $p$  values is available in our analysis scripts on OSF. 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):

Table 1  
*Means (and SD) for Control and Intentionality Composite Variables in Study 1*

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

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

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

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). Furthermore, accidents were rated as more controllable than uncontrollable acts,  $t(142) = -9.35$ ,  $p_a < .001$ , 95% CI  $[-0.95, -0.62]$ ,  $d_{mm} = 0.70$ , but not as more intentional than uncontrollable acts,  $t(142) = -2.70$ ,  $p_a = .433$ , 95% CI  $[-0.27, -0.04]$ ,  $d_{mm} = 0.14$ . 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 ( $p_a < .001$ ). 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  $[-0.20, -0.003]$ ,  $d_{mm} = 0.11$ , and intentionality,  $t(142) = -0.49$ ,  $p_a = 1$ , 95% CI  $[-0.12, 0.07]$ ,  $d_{mm} = 0.03$ . All mental state categories were rated as more intentional than accidents ( $p_a < .001$ ), and more controllable than uncontrollable behaviors ( $p_a < .001$ ). Thus, in general, mental states were judged as neither fully controllable or uncontrollable, nor fully intentional or unintentional.<sup>2</sup>

**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 (see supplemental material Tables S1 and S2 for nonadjacent category results). For both control and intentionality, emotions were rated lower than desires, control:  $t(142) = -3.67$ ,  $p_a = .019$ , 95% CI  $[-0.50, -0.15]$ ,  $d_{mm} = 0.27$ ; intentionality:  $t(142) = -8.77$ ,  $p_a < .001$ , 95% CI  $[-0.97, -0.61]$ ,  $d_{mm} = 0.65$ , desires were rated lower than beliefs, control:  $t(142) = -6.40$ ,  $p_a < .001$ , 95% CI  $[-0.67, -0.36]$ ,  $d_{mm} = 0.45$ ; intentionality:  $t(142) = -5.15$ ,  $p_a < .001$ , 95% CI  $[-0.54, -0.24]$ ,  $d_{mm} = 0.33$ , beliefs did not differ from evaluations, control:  $t(142) = -0.72$ ,  $p_a = 1$ , 95% CI  $[-0.23, 0.11]$ ,  $d_{mm} = 0.05$ ; intentionality:  $t(142) = -1.54$ ,  $p_a = 1$ , 95% CI  $[-0.29, 0.04]$ ,  $d_{mm} = 0.11$ , and evaluations were rated lower than deliberations, control:  $t(142) = -4.38$ ,  $p_a = .001$ , 95% CI  $[-0.56, -0.21]$ ,  $d_{mm} = 0.36$ ; intentionality:  $t(142) = -4.44$ ,  $p_a = .001$ , 95% CI  $[-0.61, -0.23]$ ,  $d_{mm} = 0.39$ . Deliberations were rated no differently than imaginings, control:  $t(142) = -0.84$ ,  $p_a = 1$ , 95% CI  $[-0.23, 0.09]$ ,  $d_{mm} = 0.06$ ; intentionality:  $t(142) = -1.31$ ,  $p_a = 1$ , 95% CI  $[-0.29, 0.06]$ ,  $d_{mm} = 0.11$ , and imaginings were rated lower than intentions, control:  $t(142) = -9.11$ ,  $p_a < .001$ , 95% CI  $[-1.01, -0.65]$ ,  $d_{mm} = 0.77$ ; intentionality:  $t(142) = -8.26$ ,  $p_a < .001$ , 95% CI  $[-1.01, -0.62]$ ,  $d_{mm} = 0.76$ . All nonadjacent categories (e.g., imaginings-beliefs) were rated significantly different from one another on both control and intentionality.

**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 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:  $b = 0.17$ ,  $SE = 0.03$ ,  $t = 6.33$ ,  $p < .001$ ). See supplemental material Tables S3 and S4 for in-depth comparisons and other model outputs.

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. These results are robust to different model specifications (see supplemental materials Tables S5 and S6 and Figure S1).

## Discussion

The purpose of Study 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

<sup>2</sup> It is possible that average ratings of moderate control and intentionality reflect bimodal distributions of judgments. However, examination of the distribution of ratings rules this out: The modal response for each mental state category (except for intentions) was in the middle of the scale (3–5, on 1–7 scales).

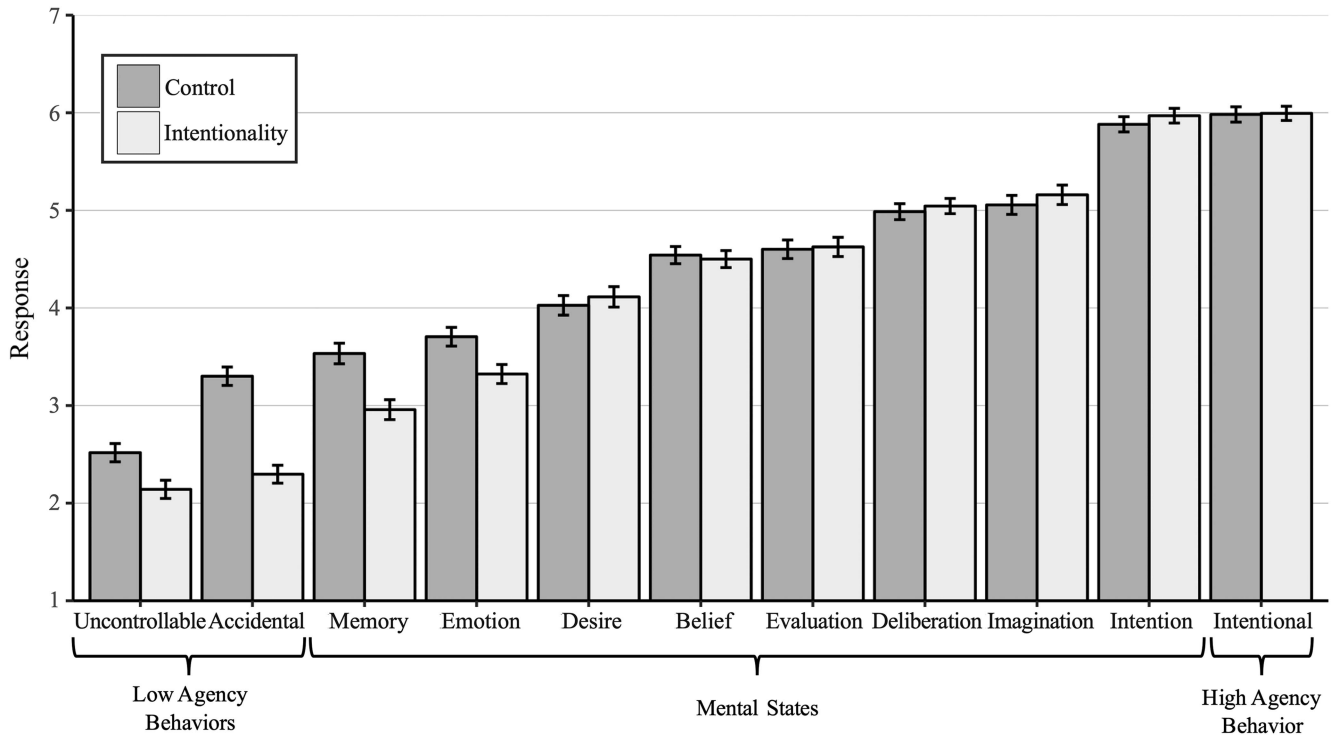


Figure 1. Means (and standard errors) of control and intentionality ratings in Study 1.

seen as willfully controllable (e.g., Adams, 1985; D'Andrade, 1987).

Study 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. In turn, consistent with D'Andrade's (1987) theorizing, beliefs were seen as more controllable than desires and emotions. Other findings, however, were inconsistent with past theorizing—for instance, desires were seen as more controllable than emotions (cf. D'Andrade, 1987). And, although beliefs and evaluations were judged the most controllable and intentional, they were not seen as completely voluntary (cf. Turri et al., 2018).

Study 1 has two notable shortcomings. First, although the mental states we tested were mundane, and likely familiar to our college student subjects (e.g., “She believed she did well on the exam”), they were presented without any surrounding context. The absence of relevant context might have either increased or decreased ratings of control and intentionality. For instance, knowledge of the context in which a mental state occurs might heighten people's appreciation of the typical constraints operating on people's minds, including different environmental stimuli, and other agents' behavior (e.g., in the example above, the agent may have experienced mastery over all of the exam items, leaving little room to believe that she had not done well). In this sense, the absence of surrounding context might have inflated ratings of intentionality and control in the present study. However, it is also possible that providing more surrounding context would bolster the intuition

that the target agent chose a particular mental state and could have chosen otherwise. Second, the content of the mental states was not held constant across mental state type. 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.

The remaining studies were designed to address both of these limitations. Study 2 addresses the first limitation (lack of context) by calling for subjects to make judgments about specific mental states experienced by someone they knew well, for which they understood the surrounding context very well. Study 3 addresses the second limitation (systematically varying content) by using entirely abstract stimuli and thereby removing any specific content. Studies 4–6 address both limitations at once by using carefully controlled stimuli that provide surrounding context while keeping content as constant as possible across mental state categories.

## Study 2

Data from Study 2 were obtained from an autobiographical recall task conducted by Cusimano and Goodwin (2017a) in which subjects recalled a recent occasion in which they were bothered by a close other's attitude. The procedure differed from that in Study 1 in three main ways: subjects rated mental states that they themselves nominated and supplied the content for, they judged perceived control for only a single mental state rather than many, and they judged objectionable mental states experienced by someone they knew well. Because many ele-



ments of the experiment are not relevant to the topic of this paper, we discuss only the control, intentionality, blameworthiness, and character-relevance measures (see supplemental material for a list of all other measures).

## Method

**Participants.** Three hundred fifty-eight University of Pennsylvania undergraduate students ( $M_{\text{age}} = 20$ , 271 reported female) completed this experiment for course credit. Sample size was determined by the number of subjects who enrolled in the study before the end of the semester. No subjects were excluded from the analysis. Because the five conditions (see below) had slightly different sample sizes, this yielded an average power just above 40% to detect small differences ( $d = .3$ ) and above 80% to detect moderate ( $d = .5$ ) differences between those conditions.

**Procedure.** After reading the consent form, subjects were randomly assigned to one of five conditions. In each condition, subjects were asked to recall an instance in which someone they knew well did something or experienced something objectionable. In four conditions, subjects recalled another person's mental state, either an emotion, a desire, a belief, or an evaluation. In a fifth condition, subjects recalled a time that someone performed a bad action. A sample instruction for the emotion condition is below (see supplemental materials for all prompts):

Please think of a specific time that someone close to you had an emotion or feeling that you found immoral, dangerous, bad, irritating, or bothersome. This could be any time that this person felt some emotion or was in a certain mood. We will use the term "feeling/emotion" to refer to these attitudes. You may also think of a time that someone close to you didn't have a feeling/emotion, and you found this to be immoral, dangerous, bad, irritating, or bothersome.

Following previous autobiographical recall procedures (Leary et al., 1998), we asked subjects to write the initials of the target person who experienced the mental state or performed the behavior, and to provide that person's sex, age, and relationship to the self at the time of the event, as well as how long ago it occurred.

Subjects then provided the details of the event in response to several open-ended prompts. We provided separate areas for them to provide (a) a description of the emotion, belief, desire, evaluation, or action, (b) how they learned about it, (c) why they thought the target person had this mental state (or performed this behavior), (d), what they found objectionable/bad/offensive about it, and (e) whether the target person ever acted on the basis of this mental state and, if so, what the effects were. After providing these open-ended responses, subjects responded to several questions about the mental state/behavior. Using 7-point rating scales, they indicated how (f) unusual, (g) bad, (h) dangerous, and (i) bothersome they considered the mental state/behavior to be.

Then, on a separate page, subjects indicated how much *Control* or *Choice* the agent had over the mental state by indicating their agreement with the statements, "This person had control over whether or not he/she had this [desire/emotion/belief/attitude]" and "This person made a deliberate choice to have this [desire/emotion/belief/attitude]" on a 7-point rating scale (1: *completely agree*, 7: *completely disagree*). These measures were our primary dependent variables. Next, subjects indicated how blameworthy the target person was for holding the mental state ("How blameworthy

was this person for having this [desire/emotion/belief/attitude]?" 1: *not blameworthy at all*, 7: *extremely blameworthy*), and how much it impacted their own (i.e., the subjects') impressions of his or her character ("How much did this [desire/emotion/belief/attitude] negatively influence your impression of this person's character?" 1: *did not negatively influence my impression at all*, 7: *negatively influenced my impression to a great degree*). Subjects then provided information about their behavioral responses to the target person (such as avoiding or punishing), which we report elsewhere, and then, lastly, filled out a demographics questionnaire. The order of these questions was fixed across all subjects, in the order presented here.

## Results

To check whether subjects followed the instructions correctly, two coders (the first author as well as one person blind to the purpose of the study) rated each of the 358 descriptions and categorized them as either emotions, beliefs, evaluations, desires, or observable actions. The reliability of these ratings was high (Cohen's  $\kappa = 0.78$ ), and coders arrived at a final coding through discussion. Subjects appeared to understand the instructions of the task, as there was good agreement between a priori condition designation and the raters' agreed upon codings ( $\kappa = 0.70$ ). All analyses below were carried out using the raters' codings.<sup>3</sup> See Appendix B for examples of subject-provided mental state contents.

Subjects' choice and control ratings were highly (though not perfectly) correlated, both overall,  $r(282) = 0.66$ ,  $p < .001$ , 95% CI [0.59, 0.72], as well as within each of the mental state categories: emotions,  $r(61) = 0.66$ ,  $p < .001$ , 95% CI [0.50, 0.78], desires,  $r(79) = 0.61$ ,  $p < .001$ , 95% CI [0.45, 0.73], beliefs,  $r(76) = 0.54$ ,  $p < .001$ , 95% CI [0.36, 0.68], and evaluations,  $r(60) = 0.65$ ,  $p < .001$ , 95% CI [0.48, 0.78]. To examine differences between each mental state category, we conducted a series of independent  $t$  tests following the "adjacent categories" procedure from Study 1 (correcting for multiple comparisons). Overall, the ordering of relative controllability was similar to that observed in Study 1 (see Figure 2). Emotions ( $M = 3.75$ ,  $SD = 1.82$ ) and desires ( $M = 4.46$ ,  $SD = 1.70$ ) were not significantly different after correcting for multiple comparisons,  $t(128.79) = -2.39$ ,  $p_a = .055$ , 95% CI [-1.30, -0.12],  $d = 0.40$ . However, desires were rated as less chosen than beliefs ( $M = 5.19$ ,  $SD = 1.53$ ),  $t(156.25) = -2.87$ ,  $p_a = .024$ , 95% CI [-1.24, -0.23],  $d = 0.45$ , which were rated no differently from evaluations ( $M = 5.44$ ,  $SD = 1.52$ ),  $t(131.25) = -0.94$ ,  $p_a = .35$ , 95% CI [-0.76, 0.27],  $d = 0.16$ . All comparisons between nonadjacent categories were significant (see Figure 2, below, and supplemental material Table S8).

We observed the same pattern for ratings of control. Although emotions ( $M = 4.32$ ,  $SD = 1.69$ ) and desires ( $M = 4.72$ ,  $SD = 1.69$ ) were not rated as significantly different,  $t(133.3) = -1.403$ ,  $p_a = .254$ , 95% CI [-0.96, 0.16],  $d = 0.24$ , desires were rated lower than beliefs ( $M = 5.36$ ,  $SD = 1.46$ ),  $t(155.18) = -2.57$ ,  $p_a = .044$ , 95% CI [-1.14, -0.15],  $d = 0.41$ . Beliefs and evaluations ( $M = 5.73$ ,  $SD = 1.36$ ) were not rated differently,  $t(134.59) = -1.54$ ,  $p_a = .254$ , 95% CI [-0.84, 0.11],  $d = 0.26$ .

<sup>3</sup> All analyses from Study 2 yield the same conclusions when analyzing mental states based on condition assignment and not the raters' recodings.

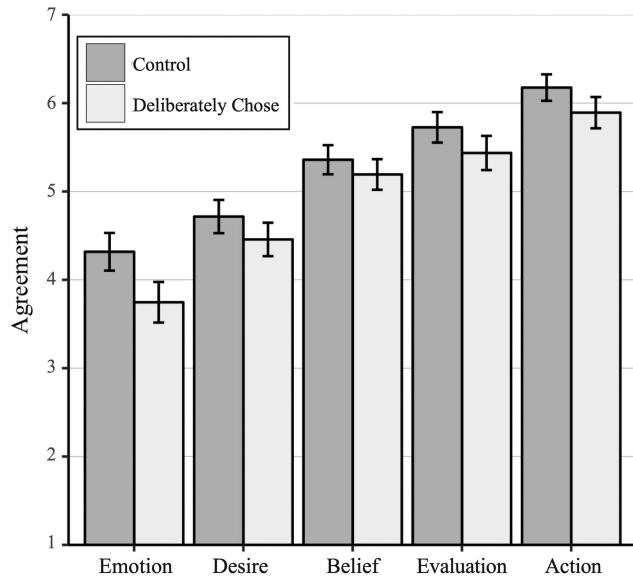


Figure 2. Means (and standard errors) of control and deliberate choice ratings in Study 2.

All comparisons between nonadjacent categories were significantly different  $p_a s < .002$ .

Also consistent with Study 1, emotions, desires, and beliefs were all rated as less chosen and less controlled than behaviors (choice:  $M = 5.89$ ,  $SD = 1.52$ ; control:  $M = 6.18$ ,  $SD = 1.29$ ;  $p_a s < .03$ ), though, interestingly, evaluations were not (choice:  $p_a = .168$ ; control:  $p_a = .152$ ). This may have been because the instructions did not specifically ask for intentional bad behaviors. Indeed, 10% of the time subjects described behaviors that were not intentional, that is, they described behaviors that they themselves rated as lower than 4 on the 1–7 intentional scale. When we removed these cases, we found that evaluations were rated as less chosen,  $t(94.21) = -4.09$ ,  $p_a < .001$ , 95% CI  $[-1.33, -0.46]$ ,  $d = 0.74$ , and less controlled,  $t(110.36) = -3.32$ ,  $p_a = .001$ , 95% CI  $[-1.12, -0.28]$ ,  $d = 0.59$ , than the behavioral controls.

Control and intentionality judgments correlated with judgments of blameworthiness, control:  $r(282) = 0.45$ ,  $p < .001$ , 95% CI  $[0.35, 0.54]$ ; intentional choice:  $r(282) = 0.42$ ,  $p < .001$ , 95% CI  $[0.32, 0.51]$ , and character-relevance, control:  $r(282) = 0.25$ ,  $p < .001$ , 95% CI  $[0.13, 0.35]$ ; intentional choice:  $r(282) = 0.19$ ,  $p = .002$ , 95% CI  $[0.07, 0.30]$ . Control correlated more strongly with blameworthiness than with character-relevance (choice:  $z = -6.9$ ,  $p < .001$ ; control:  $z = -7.89$ ,  $p < .001$ ; using Steiger's test). Blameworthiness and character-relevance were also correlated,  $r(282) = 0.60$ ,  $p < .001$ , 95% CI  $[0.52, 0.67]$ . These patterns replicated when the data were analyzed within mental state category as well: targets were consistently viewed as less blameworthy for mental states that were less controllable or intended, whereas the relationship between control (and choice) and character-relevance was weaker; see supplemental material Table S9).

## Discussion

Study 2 replicated the main results from Study 1, despite several notable differences in methodology. Overall ratings of control and

intentionality for mental states were similar to those obtained in Study 1, falling at or above the midpoint of the rating scales, though generally below the ratings for intentional behaviors (see also supplemental materials Table S13). (There was no “uncontrollable behavior” condition against which to compare such ratings in this study.) Similar to Study 1, we observed an increase in perceived agency across emotions, desires, beliefs, and evaluations (though with less differentiation between emotions and desires). These results generalize the Study 1 findings to real-world situations, and provide assurance that the lack of context in Study 1 did not systematically bias the results.

However, one limitation of both of the first two studies is that the specific contents of the mental states across categories (beliefs, desires, emotions, etc.) were not held constant. This leaves open the possibility that the differences we observed in overall perceptions of control were driven by some systematic differences in content, rather than by the nature of the mental states themselves. One strategy for addressing this is to remove content entirely, and to rely on entirely abstract representations of mental states. Another strategy is to provide content, but to hold it constant as far as possible. Study 3 deployed the first of these strategies, and Studies 4–6 deployed the second.

## Study 3

### Method

**Participants.** Four hundred ( $M_{age} = 35$ , 182 reported female, one unreported) people were recruited from Amazon's Mechanical Turk worker pool. Sample size was determined prior to data collection. No subjects were excluded. This yielded greater than 80% power to detect small ( $d = .3$ ) differences between conditions.

**Mental states.** We included 80 different mental state terms consisting of 20 belief states, 10 desire states, 10 evaluative attitudes, 10 intention terms, 10 deliberation terms, five imagination terms (imaginings),<sup>4</sup> 10 emotions, and five memory events. We also included two additional perceptual states (*see* and *hear*), though we omitted these from our analyses because they were ambiguous between intentionally and unintentionally attending to something. We also included the 15 observable behaviors from Study 1, including 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*). See Appendix C for a full list of mental state terms used in this study.

We created a series of generic statements using a nondescript subject noun phrase (“He” or “She”) paired with a verb phrase denoting each of the mental states or physical behaviors above. Each statement conformed to the following rules:

1. Emotions contained just the subject noun phrase, “felt,” and the target emotion (e.g., *She felt happy; He felt depressed; She felt embarrassed*).

<sup>4</sup> Because of experimenter error, one of the imagination terms was left out of one of the lists. Thus, we only have data for four imagine states (see Appendix C).

2. Beliefs included the phrase “something was true” (e.g., *He had the impression that something was true*).
3. Intention states included the phrase “to do something” (e.g., *She intended to do something*).
4. All other mental state categories (desires, evaluative attitudes, deliberations, imaginings, memory events, and perceptual states) took the noun “something” as the object of the verb (e.g., *He desired something; He loved something; He deliberated about something; He imagined something; He remembered something; He saw something*).
5. Uncontrollable behaviors had no object (e.g., *She shivered*).
6. All other behavior categories took the nondescript noun “something” as the direct object of the verb (e.g., *She said something; She slipped on something*).

**Procedure.** At the beginning of the experiment, subjects were randomly assigned to one of four between-subjects judgment conditions. The first three conditions probed for different judgments of control, as follows: Control: “To what extent did he/she have control over this?”; Intentionality: “How intentional do you think this was?”; Preventability: “To what extent could she have prevented this?” The last condition called for a social/moral judgment of Responsibility: “How responsible do you think she was for this?” Subjects responded to the same question for all items.

To reduce the length of the study, and to present all subjects with a similar balance of mental and behavioral items, the 80 mental states were evenly distributed across five lists containing 16 mental states each, such that each list was equivalent in terms of the number of mental states from each mental state category. The number from each category was determined in proportion to the total number of mental state terms from each category across the experiment (i.e., we selected four belief verbs, two desire verbs, and so on). Each of the five lists had a male and female version in which, for every trial, the nondescript agent was either “He” or “She.” All 15 observable behaviors and two perceptual states were used in all five lists to create constant reference categories across all subjects. This procedure resulted in 33 items per list. At the start of the experiment, subjects were randomly assigned to one of 40 conditions: either a male or female version of one of the five stimulus lists, crossed with one of the four control or responsibility questions. Each trial was presented on a separate screen in a new random order for each subject. At the end of the study, subjects were asked to report their age, sex, whether English was their first language and, if not, for how many years they had been speaking it. No other measures were collected.

## Results

We combined data across male and female versions of each list. We then calculated the item-level means for each of the eight mental state categories and three behavior categories. These means were calculated separately for each of the five item lists (recall that each of the five lists had the same mental state and behavior categories, but different individual items instantiating those categories). The means were also calculated separately for each of the

four judgment conditions (control, intentionality, preventability, responsibility). Within each judgment type, we then examined the reliability of the ratings of the 11 categories across the five lists. This revealed that ratings of control ( $\alpha = .98$ ), intentionality ( $\alpha = .98$ ), preventability ( $\alpha = .95$ ), and responsibility ( $\alpha = .99$ ) were all highly reliable. Because of this, we aggregated responses across each of the five lists into a single dataset, which resulted in 99–101 subjects means per judgment condition. Because of experimenter error, one of the stimulus lists did not include an imagination item, and so 20% of our sample had no value for the imagination condition. We did not analyze subjects’ responses to the perception items. See Table 2 for means and standard deviations for our three control measures (see supplemental material Tables S10–S12 for detailed output).

Within the control, intentionality, and preventability conditions, we conducted a series of dependent-sample *t* tests on subjects’ mean ratings of each category pair (55 pairs total), correcting for multiple comparisons. As expected, accidents (e.g., “She slipped on something”) were rated as more controllable than uncontrollable behaviors (e.g., “She shivered”),  $t(98) = 12.74$ ,  $p_a = .003$ . Accidents were also rated as more intentional than uncontrollable behaviors,  $t(99) = 3.56$ ,  $p_a = .001$ , but this difference was smaller than that for controllability ratings,  $F(1, 197) = 74.7$ ,  $p < .001$ , confirming that subjects were using these measures as expected.

Replicating results from Study 1, all mental states were viewed as more intentional than uncontrollable behaviors and accidents ( $ts > 7.10$ ;  $p_{as} < .001$ ). Every mental state category was also judged as more controllable than uncontrollable behaviors ( $ts > 11.37$ ;  $p_{as} < .001$ ), and all mental states except for emotions,  $t(98) = 1.02$ ,  $p_a = .931$ , were judged as more controllable than accidents ( $ts > 4.86$ ;  $p_{as} < .001$ ). Additionally, most mental states were seen as less controllable ( $ts \leq -3.58$ ,  $p_{as} \leq .005$ ) and intentional ( $ts \leq -3.65$ ,  $p_{as} \leq .003$ ) than intentional acts. The only exception was the category of intentions, which did not differ from intentional acts in control,  $t(98) = 0.92$ ,  $p_a = .931$ , or intentionality,  $t(99) = 1.32$ ,  $p_a = .392$ .

Preventability, which we measured for the first time in this study, exhibited a similar pattern, but with a few exceptions. First, as Table 3 shows, accidents were rated considerably higher on preventability than controllability and intentionality. As a conse-

Table 2  
Overall Means (and Standard Deviations) by Behavior/Mental State Category Across Control Measures in Study 3

Behavior	Control	Intentionality	Preventability
Uncontrollable act	2.07 (1.20) <sup>a</sup>	1.79 (1.21) <sup>a</sup>	2.60 (1.70) <sup>a</sup>
Accidental act	3.48 (1.62) <sup>b</sup>	2.05 (1.38) <sup>b</sup>	4.62 (1.72) <sup>c,d</sup>
Emotion	3.63 (1.59) <sup>b</sup>	2.90 (1.50) <sup>c</sup>	3.36 (1.78) <sup>b</sup>
Memory	4.43 (1.81) <sup>c</sup>	4.05 (1.83) <sup>d</sup>	3.76 (1.88) <sup>b</sup>
Desire	4.74 (1.59) <sup>d</sup>	4.51 (1.86) <sup>c</sup>	4.34 (1.92) <sup>c</sup>
Evaluation	5.19 (1.47) <sup>c</sup>	5.14 (1.39) <sup>f</sup>	4.33 (1.94) <sup>c</sup>
Belief	5.29 (1.47) <sup>c</sup>	5.11 (1.49) <sup>f</sup>	4.42 (1.84) <sup>c</sup>
Deliberation	5.72 (1.27) <sup>f</sup>	5.59 (1.27) <sup>g</sup>	5.01 (1.69) <sup>d</sup>
Imagining	5.91 (1.19) <sup>f</sup>	5.74 (1.05) <sup>g</sup>	4.78 (1.90) <sup>c,d</sup>
Intention	6.23 (1.07) <sup>g</sup>	6.11 (1.26) <sup>h</sup>	5.62 (1.61) <sup>c</sup>
Intentional act	6.30 (1.01) <sup>g</sup>	6.23 (1.07) <sup>h</sup>	6.03 (1.49) <sup>f</sup>

*Note.* Order arranged by average Control rating. Within each column, cells that share a superscript are not significantly different from one another. Response scales ranged from 1–7.

Table 3  
By-Item Correlations (*R* and 95% *CI*) Between Each of the Four  
Dependent Measures in Study 3

Variable	Control	Intentionality	Preventability
Control	—		
Intentionality	.89 [.83, .93]	—	
Preventability	.73 [.61, .82]	.67 [.53, .78]	—
Responsibility	.85 [.77, .91]	.89 [.83, .93]	.67 [.53, .78]

Note.  $df = 77$ . All  $p$  values were  $< .001$ .

quence, among the mental states, only intentions were rated as more preventable than accidents,  $t(100) = 5.72$ ,  $p_a < .001$ ), whereas several mental states were judged as less preventable, including emotions and memories ( $ts \leq -4.34$ ,  $p_{as} < .002$ ). Desires, beliefs, evaluations, imaginings, and deliberations were not judged as significantly more or less preventable than accidents ( $p_{as} \geq .736$ ). However, all mental states were seen as more preventable than uncontrollable behaviors ( $p_{as} < .001$ ), and less preventable than intentional acts ( $p_{as} \leq .012$ ).

Comparing the mental state categories, we observed a pattern of increasing control and intentionality, similar to the one observed in Studies 1 and 2. Emotions were rated as less controllable than desires,  $t(98) = -6.75$ ,  $p_a < .001$ ,  $d_{rm} = .83$ , which were rated as less controllable than beliefs,  $t(98) = -4.16$ ,  $p_a = .001$ ,  $d_{rm} = .47$ , and evaluations,  $t(98) = -3.51$ ,  $p_a < .001$ ,  $d_{rm} = .37$ , which in turn were not different from each other,  $t(98) = -0.92$ ,  $p_a = .931$ ,  $d_{rm} = .09$ . Beliefs were rated as less controllable than deliberations,  $t(98) = -3.75$ ,  $p_a = .005$ ,  $d_{rm} = .40$ , as were evaluations  $t(98) = -4.36$ ,  $p_a < .001$ ,  $d_{rm} = .47$ . Evaluations were rated as less controllable than imaginings,  $t(80) = -5.08$ ,  $p_a < .001$ ,  $d_{rm} = .57$ . Deliberations and imaginings were not significantly different from each other,  $t(80) = 1.66$ ,  $p_a = .443$ ,  $d_{rm} = .19$ . Deliberations were rated as less controllable than intentions,  $t(98) = -4.89$ ,  $p_a < .001$ ,  $d_{rm} = .51$ , but imaginings were rated no less controllable than intentions,  $t(80) = -1.94$ ,  $p_a = .338$ ,  $d_{rm} = .25$ . This order was perfectly replicated for intentionality judgments, despite the fact that these ratings were made by different subjects (see supplemental material Table S11).

There was less differentiation between mental states in terms of preventability judgments (see Table 3 and supplemental material Table S12). Emotions were rated as less preventable than desires, evaluations, beliefs, deliberations, and intentions ( $ts > 5.79$ ,  $p_{as} < .001$ ). Deliberations and intentions were judged more preventable than all other mental states ( $ts > 3.79$ ,  $p_{as} \leq .005$ ), though there were no differences between desires, evaluations, and beliefs ( $p_{as} \geq .999$ ).

Finally, we analyzed the relationship between each of our three measures of control and responsibility. We calculated the average rating on each of the four control measures for each of the 80 mental state items, and then ran by-item correlations. These correlations showed that, even between subjects, all control and responsibility judgments were strongly positively related ( $ps < .001$ , see Table 3).

## Discussion

Study 3 used a more abstract method that addresses the possibility that different content across the mental state categories might

have been responsible for the differences we observed previously between them. Using this method, the primary results of the earlier studies replicated. Even when specific content was removed, we again saw an increase in perceived control and intentionality across the following mental state categories: emotions  $<$  desires  $<$  beliefs, evaluations  $<$  imaginings, deliberations  $<$  intentions. Whereas the control and intentionality ratings clearly replicated findings from Studies 1 and 2, mental state types were not well distinguished by our new measure of control, preventability. It may be that the question about preventability draws in consideration of factors that occurred (or could have occurred) well prior to the mental state in question, whereas intentionality and control measures focus more on in-the-moment decisions and choices. Finally, we observed strong correlations between measures of agency (control, intentionality, and preventability) and responsibility, replicating findings from Studies 1 and 2. In this case, however, ratings for each control construct were made between subjects, and so these patterns were unlikely to have stemmed from spill-over effects, or from lay theories about how the various measures ought to relate to one another.

The use of abstract stimuli in this experiment has its advantages, but it leaves open certain questions as well. By removing specific content from the mental states, we gain some assurance that different contents did not drive the earlier results. Yet, the possibility still remains that even these abstract statements tended to make accessible certain kinds of contents that varied systematically. In other words, it is possible that subjects “filled in” in the abstract statements with the most readily available contents, and that they did so in some kind of systematic way (e.g., based on prior associations with each term). For instance, it is possible that subjects who encountered, say, the statement: “He craved something,” tended to think of a particularly strong desire for a food item (e.g., chocolate). In contrast, when encountering the statement: “He believed that something was true,” it is possible that many subjects thought of, say, belief in the existence of God. In this way, the use of abstract statements does not completely shut the door to an alternative explanation based on content. Accordingly, in Studies 4 and 5, we implemented an even tighter control procedure, by stipulating the contents of each mental state in a manner that held mental state content as constant as possible.

## Interim Discussion

Studies 1–3 converge on the following findings: many different mental states are seen as highly controllable, though not as controllable or intentional as intentional behaviors. The controllability of mental states predicts important social judgments about them, including responsibility and blame. These findings replicated across different measures of control, and when either mental state category (Study 2) or judgment type (Study 3), varied between subjects.

We also observed that not all mental states were viewed as equally controllable. Across all three studies, emotions were seen as less controllable than desires (this difference was significant in Studies 1 and 3, and approached significance in Study 2). Desires, in turn, were seen as less controllable than beliefs and evaluations, which were viewed equivalently. Beliefs and evaluations were seen as less controllable than deliberations, imaginings, and intentions.

What explains these differences? One possibility, which is suggested by the philosophical and anthropological work reviewed in the Introduction, is that the controllability people attribute to various mental state categories derives from properties that are intrinsic to each category. For instance, perhaps people's *concept* of an intention comprises a high degree of controllability (over the holding of the intention itself), whereas their concept of an emotion comprises a lack of controllability (e.g., D'Andrade, 1987; Sabini & Silver, 1998). Another possibility is that the differences we have observed have little to do with intrinsic (or conceptual) properties of the mental states but, rather, arise from other, extrinsic sources of variation—principally, the variable content of the mental states. As previously noted, perhaps people associate different mental state categories with systematically different contents, and it is these content differences that drive the effects. To continue our earlier example, perhaps people judge that a person has less control over their mind when they are interacting with chocolate (e.g., wanting it, thinking about it, etc.) than with other kinds of stimuli, and that the propensity for a mental state to have chocolate as its object is greater for desires (say) than it is for other kinds of mental states. A related possibility is that different mental states are associated with systematically different contexts, which also feeds into judgments of control—for instance, perhaps people think of interactions with chocolate occurring in social settings, which impose greater constraints on mental state control than do more solitary settings (which might in turn be associated with other belief contents). Although the abstractness of Study 3 was designed to preclude this possibility, it cannot do so entirely, and so the studies we have conducted thus far cannot adjudicate between these possibilities.

To resolve this, in Studies 4–6 we constrained the background scenario and specific content of the mental states to be extremely similar, and asked whether the differences between emotions, desires, beliefs, evaluations, and thoughts would replicate even when content is tightly controlled in this way. Accordingly, for each scenario, we attempted to create items that varied only in the mental state described, but which otherwise indicated very similar overall attitudes, for instance: “wanting X,” “believing that X would be good,” “being excited that X,” “wanting to X,” and “thinking that one should X.” If ratings of control differ in this design, then this would suggest that something intrinsic to each mental state drives its judged controllability. If ratings do not differ, however, then this would suggest that content (or context) differences across each mental state drove the differences observed in Studies 1–3.

### Study 4

In Study 4, in addition to controlling the background context and the content of each mental state as carefully as possible, we also examined three distinct measures of control using separate samples: Study 4a investigated attributions of *intentional choice*, Study 4b directly investigated judgments of *control*, and Study 4c investigated a new means of assessing control—the *ability to choose to stop* thinking, feeling, or wanting something once it has started. Based on the results from Studies 1–3, we predicted that emotions, beliefs, desires, and deliberative thoughts would all be seen as more controllable than uncontrollable behaviors, but less controllable than fully intentional behaviors, and that there would be an

increase in controllability across emotions, desires, beliefs, and finally, deliberative thoughts.

### Method

**Power analysis.** Sample size was determined prior to data collection on the basis of a simulation power analysis (see supplemental materials, p. S12 for additional detail). This indicated that a sample size of  $N = 150$  per DV provided greater than 98% power to detect a small absolute difference between conditions ( $b = .3$ ). Estimates were based on mean differences and error estimates obtained from Study 1, and were adjusted in light of the fact we planned to recruit from Internet samples and therefore anticipated more noise/random error. Because of software limitations, we recruited just shy of 150 subjects for each DV (see below).

**Participants.** Four hundred thirty-six people were recruited from Amazon's Mechanical Turk to participate in Study 4. One hundred forty-six individuals (65 female;  $M_{\text{age}} = 35$ ) participated in Study 4a (“Choice”), 145 (66 female, 1 unreported;  $M_{\text{age}} = 35$ ) participated in Study 4b (“Control”), and the remaining 145 (60 female;  $M_{\text{age}} = 34$ ) participated in Study 4c (“Choose to stop”). Subjects were randomly assigned between Studies 4b and 4c. Data collection for Study 4a occurred a couple of days prior. No subjects participated in more than one study or were excluded from any data analyses. These studies (including their sample sizes) were preregistered on AsPredicted.

**Stimuli.** We reduced the range of mental states from the earlier studies to include only four categories: emotions, desires, beliefs, and thoughts (which encompassed deliberating, imagining, and thinking about something, among others). To generate contexts, we constructed 30 scenarios describing someone in an ordinary and believable situation (such as repairing a bike, photographing a wedding, walking down the street, and so on). Unlike the contents used in Study 1, each scenario provided context about the person and the situation leading up to the mental state or observable behavior. Our primary manipulation was the last sentence in the scenario, which described either (a) an uncontrollable behavioral reaction, (b) an emotion, (c) a desire, (d) a belief, (e) thinking about or ruminating on some idea, or (f) an intentional action. Below is one of the 30 scenarios with each of the six conditions:

Katy is nearing the end of her third year in college. She's studying chemistry and biology in order to eventually apply to medical school. Any low grade will hurt her chances at getting into the top medical schools. Today, however, she struggled through the final exam in her chemistry class. She did not complete it in time and had to guess on the entire last page of questions.

Walking out of the exam, Katy . . .

1. begins shivering in the cold. (uncontrollable behavioral reaction)
2. feels angry at her professor. (emotion)
3. wants to leave her professor a poor course evaluation. (desire)
4. believes that her professor deserves a poor course evaluation. (belief)
5. thinks about leaving her professor a poor course evaluation. (thought)

- 6. fills out a negative course evaluation on her phone. (intentional behavior)

As this example illustrates, the context prior to the manipulation was held constant, as was the attitudinal content of each mental state (e.g., in the item above, a negative and retaliatory attitude toward the professor is conveyed in each case). The same attitude was also conveyed by the intentional behavior (leaving a poor course review). There was no such content for the uncontrollable reaction foils (shivering, sneezing, coughing, etc.). We varied the kind of emotion experienced by the agent in the scenario: across the 30 sets, the emotion condition featured the agent feeling either *angry*, *sad*, *afraid*, *excited*, or *pleased*. Given 30 scenarios, each of which yielded six endings, there were 180 items in the whole experiment. See the supplemental materials for full text for every scenario and condition.

**Design and task.** The items were distributed across six lists (of 30 scenarios each) using a Latin square design. These lists were constructed such that, for each scenario, each subject responded to exactly one of the six mental state (or behavior) conditions. Across all 30 scenarios, this meant that each subject responded to exactly five instances of each of the six mental state (or behavior) conditions. Furthermore, every condition for every item appeared once in one of the six lists. Finally, we balanced the distribution of emotion trials so that each of the five different emotions appeared within each list.

**Dependent measures.** In Study 4a, the main dependent variable was to indicate whether the agent chose the particular mental state he or she had at the end of the story. Subjects responded on a rating scale ranging from 1 (*definitely did not choose*) to 7 (*definitely did choose*). As in Study 1, the full content of each item was included in each question and was italicized (e.g., “Did Katy choose to *feel angry at her professor?*”). Study 4b measured perceptions of how much control the agent had over whether he or she had the particular mental state (or over the behavior). For instance, in the Katy vignette above, subjects were asked “How much control did Katy have over whether she *felt angry at her professor?*” on a scale from 1 (*no control at all*) to 7 (*complete control*). Finally, Study 4c measured perceptions of the degree to which the agent could stop the particular mental state (or

behavior) once it had started. In the Katy vignette above, subjects were asked, “Can Katy choose to stop *feeling angry at her professor?*” on a scale from 1 (*definitely cannot choose*) to 7 (*definitely can choose*).

**Procedure.** At the beginning of the experiment subjects were randomly assigned to one of the six stimulus lists. Subjects were provided brief instructions that they would read 30 stories about different characters and answer a question about a behavior that the character performed. Each trial was presented on a separate page in a new random order for each subject. At the end of the study, subjects filled out a brief demographics questionnaire. No other data were collected.

**Results**

**Analysis overview.** We ran a series of LMEMs regressing ratings from every trial (30 per subject) on the six-level mental state factor. We included random intercepts for subject and scenario ratings, as well as random slopes for by-subject and by-scenario variation in the effect of condition. For each control variable, *intentionality*, *control*, and *stopping*, we ran three sets of analyses. One analysis compared the means of the four mental state conditions to the uncontrollable behavior foil (dummy coded as the reference level). Another analysis compared the means of the four mental states to the intentional act behavior foil. Finally, the third analysis compared adjacent mental states with each other following our hypothesized stepwise increase in control from emotions to desires, desires to beliefs, and beliefs to thoughts. We adjusted *p* values using the same procedure in previous studies. See Table 4 and Figure 3 for results.

**Intentional choice.** Replicating results from our earlier studies, emotions ( $M = 3.95, SD = 2.08$ ), desires ( $M = 5.48, SD = 2.08$ ), beliefs ( $M = 5.43, SD = 1.73$ ), and thoughts ( $M = 5.29, SD = 1.77$ ) were all judged significantly more chosen than uncontrollable behaviors ( $M = 1.67, SD = 1.37; p_{a,s} < .001$ ), as well as significantly less chosen than intentional acts ( $M = 6.45, SD = 1.13; p_{a,s} < .001$ ). Replicating the earlier studies, emotions were rated as less chosen than desires ( $b = 1.53, SE = 0.08, t = 19.24, p_a < .001$ ) but, contrary to expectations, there were no detectable

Table 4  
Results From Three Models Comparing Mental State Conditions With Each Other, With the Uncontrollable Behavior Foil, and With the Intentional Behavior Foil in Study 4

Model	Fixed effects	Choose (Study 4a)			Control (Study 4b)			Choose to stop (Study 4c)		
		<i>B</i> ( <i>SE</i> )	<i>t</i>	<i>p<sub>a</sub></i>	<i>B</i> ( <i>SE</i> )	<i>t</i>	<i>p<sub>a</sub></i>	<i>B</i> ( <i>SE</i> )	<i>t</i>	<i>p<sub>a</sub></i>
1	Uncontrollable vs.									
	Emotion	2.28 (.19)	12.24	<.001*	2.17 (.18)	12.41	<.001*	2.41 (.19)	13.03	<.001*
	Desire	3.82 (.16)	23.41	<.001*	2.87 (.21)	13.70	<.001*	2.71 (.16)	16.50	<.001*
	Belief	3.77 (.15)	24.59	<.001*	3.00 (.21)	14.67	<.001*	2.74 (.19)	14.54	<.001*
2	Thinking	3.62 (.16)	22.29	<.001*	2.91 (.18)	15.75	<.001*	3.05 (.18)	16.60	<.001*
	Intentional vs.									
	Emotion	-2.51 (.18)	-13.65	<.001*	-2.06 (.15)	-13.55	<.001*	-1.65 (.17)	-9.93	<.001*
	Desire	-.97 (.13)	-7.51	<.001*	-1.37 (.14)	-9.98	<.001*	-1.34 (.15)	-8.89	<.001*
3	Belief	-1.02 (.12)	-8.19	<.001*	-1.23 (.13)	-9.87	<.001*	-1.31 (.15)	-8.81	<.001*
	Thinking	-1.17 (.12)	-9.47	<.001*	-1.33 (.12)	-11.06	<.001*	-1.01 (.15)	-6.95	<.001*
	Emotion vs. Desire	1.54 (.16)	9.71	<.001*	.69 (.18)	3.94	<.001*	.31 (.15)	2.04	.082
	Desire vs. Belief	-.05 (.14)	-.32	.750	.14 (.12)	1.10	.540	.03 (.14)	.23	.815
	Belief vs. Thinking	-.15 (.10)	-1.53	.252	-.10 (.09)	-1.04	.540	.30 (.14)	2.23	.078

Note. We report adjusted *p* values.  
\* *p* < .05.

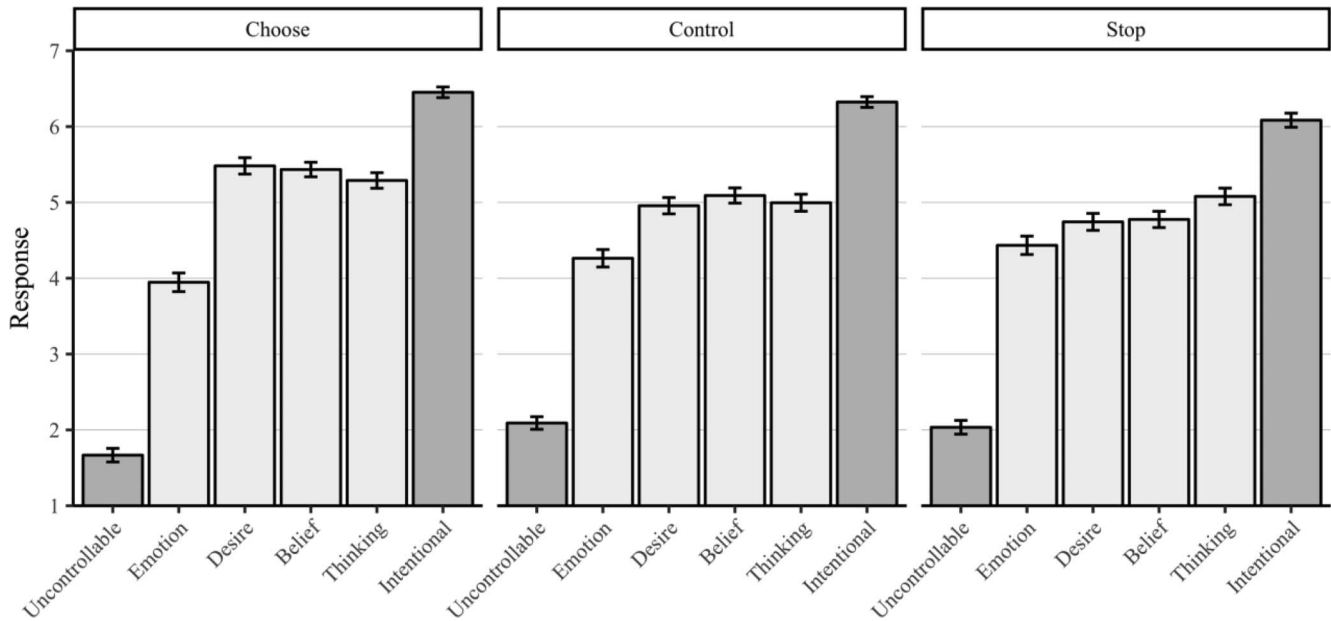


Figure 3. Condition means (and standard errors) for each control measure in Study 4. The behavior foils are shown in dark gray.

differences between desires and beliefs or between beliefs and thoughts ( $p_a s \geq .252$ ).

**Control.** Emotions ( $M = 4.26$ ,  $SD = 1.44$ ), desires ( $M = 4.96$ ,  $SD = 1.85$ ), beliefs ( $M = 5.09$ ,  $SD = 1.72$ ), and thoughts ( $M = 5.00$ ,  $SD = 1.81$ ) were all judged significantly more controllable than uncontrollable behaviors ( $M = 2.09$ ,  $SD = 1.44$ ;  $p_a s < .001$ ) and significantly less controllable than intentional acts ( $M = 6.32$ ,  $SD = 1.23$ ;  $p_a s < .001$ ). Similar to Study 4a, we observed a significant difference between emotions and desires in judged controllability ( $b = 0.69$ ,  $SE = 0.08$ ,  $t = 9.13$ ,  $p_a < .001$ ), but not between desires, beliefs, or thoughts ( $p_a s \geq .298$ ).

**Choosing to stop.** Again, emotions ( $M = 4.44$ ,  $SD = 1.99$ ), desires ( $M = 4.74$ ,  $SD = 1.98$ ), beliefs ( $M = 4.77$ ,  $SD = 1.93$ ), and thoughts ( $M = 5.08$ ,  $SD = 1.84$ ) were all judged as more stoppable than uncontrollable reactions ( $M = 2.03$ ,  $SD = 1.58$ ;  $p_s < .001$ ) and significantly less stoppable than intentional acts ( $M = 6.08$ ,  $SD = 1.54$ ;  $p_s < .001$ ). After correcting for multiple comparisons, subjects did not judge emotions to be less stoppable than desires ( $b = 0.31$ ,  $SE = 0.08$ ,  $t = 3.66$ ,  $p = 0.082$ ). Just as for choose and control judgments, beliefs were not judged to be less stoppable than thoughts ( $b = 0.31$ ,  $SE = 0.08$ ,  $t = 3.64$ ,  $p = .078$ ), nor were desires judged differently from beliefs ( $b = 0.03$ ,  $SE = 0.08$ ,  $t = 0.36$ ,  $p = .815$ ).

We also conducted a series of exploratory analyses investigating the relationship between the control measures (see Supplemental Material Table S14 and Figure S2). These analyses replicated the findings from Study 3 showing a strong positive relationship between control measures even when those measures are assessed between-subjects.

## Discussion

Across Studies 4a, 4b, and 4c, subjects rated all mental states as far more controllable than uncontrollable behaviors, no matter the

dependent variable. Thus, even when making the background context highly salient and controlling for its content, subjects still attribute to others moderate choice, control, and undoing power over their emotions, beliefs, desires, and thoughts. However, for all mental states, the degree of control attributed was less than that attributed over intentional behaviors performed on the basis of these mental states (e.g., believing that one should leave the professor a poor review was seen as less chosen, controllable, and stoppable than actually writing that review). Thus, across Studies 1–4 we have a good deal of evidence that people commonly think of mental states such as emotions, beliefs, desires, and thoughts as controllable.

Compared with Studies 1–3, Study 4 makes an important new contribution. When holding the background context constant across the different mental states, emotions were still seen as reliably less controllable than other mental states on the measures of choice and control (though not on the measure of stoppability). However, we failed to observe differences between desires and beliefs, or between beliefs and thoughts, across any of our measures of control. Judgments that a mental state could be undone (i.e., that someone could choose to stop holding the mental state) were the least differentiated of our three measures: after correcting for multiple comparisons the difference between emotions and desires, which were highly differentiated by other control measures, was only marginally significant.

One difference between this study and the earlier ones is that we used a different measure of intentionality in the present study, asking whether the protagonist chose to have each mental state (rather than asking about intentionality directly). But this feature alone cannot account for the discrepancy in the results, because the mental states were relatively undifferentiated on the control measure as well, which was very similar to that used in prior studies. When one inspects the pattern of means across these first four

studies, the major difference between Study 4 and its predecessors concerns the ratings of desires. Whereas in prior studies ratings of the controllability of desires were typically between 4 and 5 on the 7-point scale, in the present study desires were rated 5.48 on average on the intentional choice measure and 4.96 on the controllability measure.

This discrepancy may have arisen precisely because of an inherent limitation that arises when holding mental state content constant. In the illustrative example above, the intentional behavior was leaving a professor a poor course review; the corresponding desire was *wanting* to leave a professor a poor course review, the corresponding belief was *believing* that he deserves a poor review, and the corresponding thought was *thinking* about leaving such a review. In this case, all of the mental states are plausible antecedents of the same intentional behavior. The desires used in our previous experiments, however, were rarely tied to a specific behavior. For instance, in Study 1, desires were almost always more generic and less behavioral (“He desired her,” “She craved approval”), and when they were partly behavioral, they usually did not refer to a specific behavior that could be performed in the immediate context (e.g., “She desired to take a vacation,” “He desired to be outside rather than in class,” “She desired to meet her favorite singer”). Because the desires examined in Study 4 were tied to specific, intentional behaviors, it seems likely that they contained a more deliberative element than the desires presented in previous studies: wanting to leave a professor a poor course review sounds a little as though the protagonist is in fact intending to do so (Malle & Knobe, 2001). This may explain the lack of a reliable difference in ratings of desires and beliefs in the present study.

If this account is correct, then we might expect the previous differences between mental states to reemerge in a context where those states are not relevant to any action, for instance in cases in which the agent is a passive viewer of some event over which they have little to no control. We therefore designed Studies 5 and 6 to test this possibility.

### Studies 5 and 6

In Studies 5 and 6, we developed several new scenarios, all of which described the protagonist having a reaction to a set of events over which they had little to no influence. This way, unlike Study 4, the target mental states were not linked to any specific intentional action, nor could they be misconstrued as intentions. The main point of these studies was to examine whether the revised pattern of mean ratings for mental states observed in Study 4 would replicate when the mental states were no longer the antecedents of an intentional behavior.

In contrast to Study 4, we examined emotions, desires, beliefs, and evaluations. We included evaluations rather than thoughts (deliberations) because they were more relevant to the present scenarios than were deliberations, and because we were interested in whether highly similar emotions and evaluations would continue to be rated differently, as they had been in Studies 1–3. We also changed the “choose” question, reverting back to the same intentionality question used in Studies 1 and 3. In Study 5, we used immoral mental states, rather than the mundane mental states used in Study 4, which allowed us once again to examine the relationship between control and blame for mental states (as we had done in Study 2). In Study 6, we used positively valenced mental states.

## Study 5

### Method

**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 4, 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 ( $b = .30$ ).<sup>5</sup> The study (including its sample size) was preregistered on AsPredicted.

**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 D).

**Assignment of conditions.** The four mental states 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

<sup>5</sup> A simpler power analysis based on standardized mean differences of subject ratings across conditions yields an estimated power greater than 98% to detect effect sizes as small as  $d = .3$ .



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

**Analytic procedure.** We followed the same procedure as in Study 3, conducting 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 condition by scenario. In each model we regressed intentionality (or stop) on a single predictor, our four-level mental state variable, with a priori contrasts between emotion and desire, desire and belief, and belief and evaluation. See Table 5 for means and standard deviations for each of the dependent measures.

**Intentionality and stop analyses.** Replicating our prior studies, emotions were rated as less intentional than desires ( $b = 0.92$ ,

$SE = 0.13$ ,  $t = 7.02$ ,  $p < .001$ ). However, in contrast to Studies 1–4, 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 ( $b = -0.13$ ,  $SE = 0.25$ ,  $t = -0.55$ ,  $p = .584$ ).

Given these partially unexpected results, we followed up our planned analyses with further exploratory tests of the difference between emotions, desires, beliefs, and evaluations for intentionality and stop ratings, correcting for multiple comparisons. For each analysis, we removed subjects who did not respond to the control question across all four mental state conditions ( $n_{\text{intentionality}} = 6$ ,  $n_{\text{stop}} = 7$ ). Paired  $t$  tests revealed that emotions were rated as significantly less intentional than beliefs,  $t(193) = -3.54$ ,  $p_a = .002$ , 95% CI  $[-0.73, -0.21]$ ,  $d_{rm} = 0.26$ , and evaluations,  $t(193) = -4.62$ ,  $p_a < .001$ , 95% CI  $[-0.72, -0.29]$ ,  $d_{rm} = 0.27$ . However, there was no such differentiation for judgments of the ability to stop having the mental state; emotions were not rated as significantly different from either beliefs,  $t(192) = -1.89$ ,  $p_a = .419$ , 95% CI  $[-0.47, 0.01]$ ,  $d_{rm} = 0.14$ , or evaluations,  $t(192) = -0.87$ ,  $p_a = .926$ , 95% CI  $[-0.32, 0.12]$ ,  $d_{rm} = 0.06$ .

**Blameworthiness and character.** We conducted a series of four linear mixed-effect models 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. 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 ( $ps < .001$ ). 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:  $b = 0.33$ ,  $SE = 0.03$ ,  $t = 9.74$ ,  $p < .001$ ). Intentionality and stop remained significant predictors when wrongness was omitted from the models. Correlations between each of the dependent measures within each mental state condition are presented in the supplemental materials (Table S15).

Table 5  
Means (and Standard Deviations) for the Dependent Variables in Study 5 by Mental State Condition

Mental state	Intentional	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 7-point scale anchored at 1 and 7.

We next investigated whether the intentionality and stop variables were better predictors of blameworthiness judgments than character judgments using the same procedure as in Study 1. Intentionality and stop judgments were analyzed independently (see Table 6). Each LMEM included the control DV (intentionality or stop), wrongness of the attitude, attribution type (blame or character), and the interaction of control DV and attribution type. We also included random intercepts for subject and mental state category as well as random slopes for each of the predictors, including the interaction between attribution type and our control predictors. These analyses indicated that both intentionality and ability to stop judgments more strongly predicted blameworthiness judgments than character judgments (Intentionality  $\times$  Attribution type:  $b = -0.11$ ,  $SE = 0.04$ ,  $t = -2.76$ ,  $p = .006$ ; Stop  $\times$  Attribution type:  $b = -0.10$ ,  $SE = 0.05$ ,  $t = -2.17$ ,  $p = .03$ ).

### Study 6

#### Method

The design and procedures for Study 6 were almost identical to those used in Study 5, except that we presented prosocial mental states instead of morally objectionable ones, and we asked about responsibility for these positive mental states rather than blameworthiness. Additionally, our target sample size and analysis procedure were identical to those in Study 5. Our goal was to determine whether the results obtained in Study 5 generalized to nonobjectionable, prosocial mental state content.

**Participants.** Two hundred two subjects (82 female,  $M_{\text{age}} = 34.6$ ) were recruited from Amazon’s Mechanical Turk.

**Stimuli.** As in Study 5, we constructed four vignettes in which a target character learns about a state of affairs over which they have little or no control. The vignettes always ended with a positive or encouraging outcome. But, unlike in Study 5, the target always reacted in a positive way, indicating a prosocial orientation. They experienced either a positively valenced emotion (“happy” or “excited” about the good outcome), desire (“wanting” or “desiring” the good outcome), belief (“thinking” or “believing” that the good outcome is preferable), or evaluative attitude (“liking” or “loving” something about the good outcome). See below for the full text of one scenario. (See Appendix E for full text for all scenarios.)

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 but has called to tell her parents she and her long-time boyfriend just got engaged. When he hears this, James . . .

feels happy/excited that his daughter is getting married. [Emotion]

wants/desires his daughter to get married. [Desire]

thinks/believes that it is good that his daughter is getting married. [Belief]

likes/loves the fact that his daughter is getting married. [Evaluation]

The three other scenarios (with the prosocial desires presented as illustrative) involved a man learning that his mother was involved in a car accident and wanting her to survive, a civilian learning about a UN military operation designed to block murderous terrorists and wanting the mission to succeed, and a student watching a documentary about polar bears adapting to climate change, and wanting the bears to survive (for full details, see Appendix E).

Mental states were assigned to scenarios in the same way as in Study 5, with four item lists then constructed in the same way as well. Subjects were randomly assigned to one of these four lists (and to one of the two possible mental states for each scenario), with the order of presentation of items within each list randomly determined.

**Dependent measures.** Subjects answered four questions in response to each mental state. They judged the intentionality of the mental state (e.g., “Did James **intentionally** feel excited that his daughter is getting married?” 1: *definitely not intentionally*, 7: *definitely intentionally*), whether the agent could stop having the mental state (“Can James **choose to stop** feeling excited that his daughter is getting married?” 1: *definitely cannot choose to stop*, 7: *definitely can choose to stop*), the agent’s responsibility (“How **responsible** is James for feeling excited that his daughter is getting married?” 1: *not responsible at all*, 7: *extremely responsible*) and the agent’s character (“How revealing is it of James’s **character** that he feels excited that his daughter is getting married?” 1: *not revealing at all*, 7: *extremely revealing*). All ratings were made on

Table 6  
Model Outputs Comparing the Strength of the Relationship Between Control DVs (Intentionality and Stop) and Attribution Type (Blameworthiness or Character Judgments) Judgments in Study 5

Parameter	Control DV: Intentionality				Control DV: Stop			
	B	SE	t	p	B	SE	t	p
Intercept	.37	.14	2.74	.006	.56	.26	2.14	.032
Wrongness	.68	.10	6.84	.017	.71	.03	25.82	<.001
Control DV	.24	.12	2.02	.044	.16	.04	4.06	<.001
Attribution type	.73	.30	2.40	<.001	.72	.31	2.35	.019
Attribution type $\times$ Control DV	-.11	.04	-2.76	.006	-.10	.05	-2.17	.030
R <sup>2</sup>	.65				.65			

Note. All model parameters were significant ( $p < .05$ ). Of principle interest was the interaction term: Attribution type (Blameworthiness, Character)  $\times$  Control DV. A negative value indicates that the control DV was a stronger correlate of blameworthiness than character.

a 7-point rating scale ranging from 1 to 7. The order of the questions was randomly determined for each trial.

## Results

Our analysis procedures exactly mirrored those in Study 5. Emotions were rated as less intentional than desires ( $b = 0.48$ ,  $SE = 0.14$ ,  $t = 3.55$ ,  $p < .001$ ), but desires and beliefs did not significantly differ ( $b = -0.16$ ,  $SE = 0.15$ ,  $t = -1.28$ ,  $p = .31$ ), nor did beliefs and evaluations ( $b = -0.03$ ,  $SE = 0.13$ ,  $t = -0.24$ ,  $p = .81$ ). Similar to the findings from Study 5, we observed less differentiation when analyzing stop judgments. There were no reliable differences between emotions and desires ( $b = 0.3$ ,  $SE = 0.19$ ,  $t = 1.61$ ,  $p = .11$ ), desires and beliefs ( $b = 0.29$ ,  $SE = 0.21$ ,  $t = 1.36$ ,  $p = .17$ ), or beliefs and evaluations ( $b = -0.29$ ,  $SE = 0.23$ ,  $t = -1.3$ ,  $p = .19$ ). See Table 7 for means and standard deviations for each DV across condition.

Thus, the findings were highly similar to those observed in Study 5, notwithstanding the use of prosocial content. Emotions were rated as less intentional than other mental states, with less differentiation between the other mental states. And there was very little differentiation between mental states on the ability to stop measure. To confirm this picture, we followed up these analyses by conducting dependent  $t$  tests on subject's intentionality and stop judgments, as in Study 5, again correcting for multiple comparisons. Emotions were not rated as significantly less intentional than beliefs,  $t(201) = -2.47$ ,  $p_a = .071$ , 95% CI  $[-0.59, -0.07]$ ,  $d_{rm} = -0.17$ , or evaluations,  $t(201) = -2.26$ ,  $p_a = .099$ , 95% CI  $[-0.56, -0.04]$ ,  $d_{rm} = -0.15$ , though both comparisons approached significance. Contrary to what we observed in Study 5, we found that emotions were rated as less stoppable than beliefs,  $t(201) = -4.08$ ,  $p_a < .001$ , 95% CI  $[-0.87, -0.3]$ ,  $d_{rm} = -0.3$ , but no less stoppable than evaluations,  $t(201) = -2.36$ ,  $p_a = .095$ , 95% CI  $[-0.54, -0.05]$ ,  $d_{rm} = -0.15$ .

As in Study 5, we conducted a series of four linear mixed-effect models to investigate whether judgments of intentionality (or, separately, stop) predicted attributions of responsibility and character, though without wrongness as a predictor (as we did not measure it). Replicating prior studies, intentionality predicted responsibility ( $b = 0.37$ ,  $SE = 0.03$ ,  $t = 12.85$ ,  $p < .001$ ), and character judgments ( $b = 0.18$ ,  $SE = 0.03$ ,  $t = 6.38$ ,  $p < .001$ ). However, the relationship between intentionality and responsibility judgments was significantly stronger than that between intentionality and character judgments ( $b = 0.26$ ,  $SE = 0.03$ ,  $t = 8.35$ ,  $p < .001$ ). Stop judgments similarly predicted responsibility judgments ( $b = 0.23$ ,  $SE = 0.03$ ,  $t = 7.82$ ,  $p < .001$ ); however, in contrast to prior studies, stop judgments did not predict character

judgments ( $b = 0.03$ ,  $SE = 0.03$ ,  $t = 1.08$ ,  $p = .28$ ). The stop and responsibility relationship was significantly stronger than the stop and character relationship ( $b = 0.26$ ,  $SE = 0.03$ ,  $t = 8.35$ ,  $p < .001$ ). These patterns of association between each of our control judgments and each of our social judgments replicated within each of the four mental state categories considered separately (correlations for each mental state category are available in the Supplemental Material, Table S16).

## Discussion for Studies 5 and 6

Studies 5 and 6 largely replicated what we had observed in Study 4: when the content and background details for each type of mental state were held constant, many of the previously observed differences grew considerably weaker or disappeared, and occasionally reversed (e.g., in the case of the intentionality of beliefs and desires, Study 5). As in Study 4, the ability to stop having the mental state was the most undifferentiated of our control measures, revealing fewer overall differences between the mental state categories. And, unlike in Study 4, these results cannot be attributed to subjects viewing the various mental states as each precipitating the same action since, in Studies 5 and 6, the protagonists had no control over the events they learned about.

The notable exception to this general pattern was the low perceived intentionality of emotions relative to the other mental states. In Studies 4 and 5, and to some degree in Study 6, emotions were judged to be brought about less intentionally than other types of mental states. This result is consistent with the lower ratings of intentional control for emotions relative to other mental states in Studies 1–3. Replicating this difference in Studies 4–6 provides evidence that the difference between emotions and other mental states is attributable not only to differences in the prototypical contents of these mental states, but also to their differing natures (i.e., their differing intrinsic properties). The lack of differentiation in the measure of the ability to stop having the mental state indicates that the perceived difference between emotions and other mental states primarily pertains to judgments of intentionality, rather than judgments of their modifiability.

Studies 5 and 6 replicated two other primary findings. First, the pattern of relationships between the control measures and blame (Study 5), responsibility (Study 6) and character replicated the findings from Studies 1–3. Attributions of control predicted both blame and responsibility more strongly than they predicted character judgments. Second, attributions of control to mental states remained high, regardless of which control measure was used. In Study 5, the average intentionality attributed to immoral mental states was 5.15 (on a 7-point scale), and the average ability to stop them was 5.41; similarly, in Study 6, the average intentionality attributed to prosocial mental states was 4.74, and the average ability to stop them was 4.51.

## Quantitative Comparison of Context Differences

Studies 1–3 and Studies 4–6, which differed in whether we allowed the content and context of the mental states to vary, yielded qualitatively different patterns of findings. Many of the differences we observed between mental states failed to replicate when we held the content and context of those states constant. In light of our discussion above, this suggests that the differences we

Table 7  
Means (and Standard Deviations) for Dependent Variables in Study 6 by Mental State Condition

Mental state	Intentional	Stop	Responsibility	Character
Emotion	4.47 (1.92)	4.22 (1.91)	4.87 (1.63)	5.13 (1.48)
Desire	4.95 (1.72)	4.51 (1.89)	4.81 (1.77)	5.21 (1.43)
Belief	4.79 (1.84)	4.81 (1.96)	5.14 (1.59)	5.11 (1.48)
Evaluation	4.76 (1.92)	4.51 (1.90)	4.94 (1.64)	5.11 (1.48)

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

Table 8  
*Estimated Effect Sizes (d), Standard Errors (SE), and Significance Tests (z) for Context-Free and Context-Constrained Studies, With Standardized Differences (Z), for Intentionality Ratings*

State 1	State 2	Studies 1–3			Studies 4–6			Difference
		<i>d</i>	<i>SE</i>	<i>z</i>	<i>d</i>	<i>SE</i>	<i>z</i>	<i>Z</i>
Emotion	Desire	-.74	.07	-10.99*	-.51	.05	-10.36*	-2.70*
Emotion	Belief	-1.14	.08	-13.46*	-.37	.05	-8.05*	-7.97*
Emotion	Evaluation	-1.22	.08	-14.65*	-.22	.05	-4.89*	-10.56*
Desire	Belief	-.41	.05	-7.46*	.13	.04	3.08*	-7.97*
Desire	Evaluation	-.48	.05	-9.21*	.16	.05	3.19*	-8.87*
Belief	Evaluation	-.09	.06	-1.60	<.01	.05	-0.03	-1.20

Note. Evaluations were not included in Study 4.

\*  $p < .05$ .

observed in Studies 1–3 do not reflect lay people’s judgments about properties intrinsic to mental states. However, it is invalid to infer that constraining context significantly reduced the differentiation between mental state categories merely because there were no longer significant differences in the latter studies (Gelman & Stern, 2006). To test our supposition that holding context fixed across mental states attenuates the differences between them, we compared the weighted average effect sizes for differences between mental states in Studies 1–3, in which context was unconstrained, and Studies 4–6, in which context was fixed. We did this for both intentionality and controllability ratings. We focused these analyses on four mental states that were of greatest theoretical interest and that occurred most frequently across our studies: emotions, desires, beliefs, and evaluations.

We obtained standardized effect sizes ( $d$  in Study 2,  $d_{rm}$  in Studies 1, 3, 5, and 6) from the analyses reported above. To obtain comparable standardized effect size measurements from Study 4, we computed mean intentionality (or control) ratings for each mental state for each subject, and then conducted paired comparisons between the mental state conditions. We then computed weighted averages of the standardized differences for each mental state comparison for Studies 1–3 and Studies 4–6. To test whether mental state context (unconstrained vs. constrained) moderated the differences in intentionality (or controllability) judgments, we calculated the standardized difference between weighted effect sizes,  $Z$ :

$$Z = \frac{d_1 - d_2}{\sqrt{v_1 + v_2}}$$

where  $d$  is the weighted mean effect size and  $v$  is the sampling variance, calculated for studies 1 and 3–6 as follows:

$$v_d = \left( \frac{1}{n} + \frac{d^2}{2n} \right) 2(1 - r)$$

where  $r$  is the correlation between pairs of observations. Table 8 contains weighted effect size estimates for Studies 1–3 and 4–6, for intentionality judgments.

This analysis revealed that, across most comparisons, differences between mental states were significantly smaller in Studies 4–6 relative to Studies 1–3. For instance, even though emotions were still viewed as significantly less intentional than desires ( $d = 0.51$ ), beliefs ( $d = -0.37$ ), and evaluations ( $d = -0.22$ ) in Studies 4–6, the differences between emotions and beliefs ( $Z = -7.97$ ,

$p < .001$ ), and between emotions and evaluations ( $Z = -10.56$ ,  $p < .001$ ) were significantly smaller than the differences estimated in Studies 1–3. Similarly, the differences between desires and beliefs, and between desires and evaluations, which were moderate in Studies 1–3, were significantly different in Studies 4–6 when context was constrained<sup>6</sup>. Finally, we did not observe a significant effect of study type on the belief-evaluation comparison: beliefs and evaluations were not significantly different either in Studies 1–3 ( $d = -0.09$ ) or Studies 4–6 ( $d = -.001$ ).

Table 9 shows the same set of analyses for control ratings. These ratings included the composite control measure from Study 1, the measure of control over having the mental state from Studies 2–4, and the ability to stop having the mental state measure from Studies 5 and 6. We observed a pattern of significant and nonsignificant effects, and corresponding moderation, that was similar to that seen with the intentionality ratings. That is, we observed significant moderation by study type for the same five (of six possible) mental state comparisons. The consistency of this effect across both control and intentionality measures is notable given that in two studies, Studies 3 and 4, control and intentionality were assessed between subjects.

These analyses reveal that, across almost all mental state comparisons, differences in perceived intentionality and control were diminished when the context and content were held fixed. As we observed in Studies 4–6, this attenuation often resulted in the lack of any reliable difference in perceived control between the mental states. The one consistent exception to this was the reduced control and intentionality attributed to emotions compared with the other mental states. Overall, these findings suggests that considerable variation in lay judgments of the controllability and intentionality of different mental states is explained by contextual and content features that are not part of people’s underlying representations of the mental states themselves.

<sup>6</sup> Because the sign of these differences was reversed in Studies 4–6 (see Table 9), this test does not indicate whether there was any difference in absolute terms. To test whether these differences were smaller (rather than just different) in Studies 4–6, we reran our analysis investigating the absolute difference in effect sizes. This yielded the same result: the differences between desires and beliefs ( $Z = -3.80$ ,  $p < .001$ ) and between desires and evaluations ( $Z = -4.08$ ,  $p < .001$ ) were smaller once the content and context of the mental state was held constant.

Table 9  
*Estimated Effect Sizes (d), Standard Errors (SE), and Significance Tests (z) for Context-Free and Context-Constrained Studies, With Standardized Differences (Z), for Control/Stop Ratings*

State 1	State 2	Studies 1–3			Studies 4–6			Difference
		<i>d</i>	<i>SE</i>	<i>z</i>	<i>d</i>	<i>SE</i>	<i>z</i>	<i>Z</i>
Emotion	Desire	-.38	.06	-6.01*	-.17	.04	-4.02*	-2.74*
Emotion	Belief	-.82	.08	-10.84*	-.23	.04	-5.50*	-6.84*
Emotion	Evaluation	-.89	.08	-11.71*	-.11	.05	-2.30*	-8.72*
Desire	Belief	-.48	.06	-8.02*	-.06	.04	-1.57	-6.84*
Desire	Evaluation	-.49	.05	-9.04*	.03	.05	0.58	-6.93*
Belief	Evaluation	-.03	.06	-0.50	.11	.05	2.37*	-1.90

Note. Evaluations were not included in Study 4.

\*  $p < .05$ .

### General Discussion

Our aim in this paper was to examine a question that has hitherto only been addressed obliquely in existing literature, namely, to what extent people attribute to others control over their mental states. We wanted to estimate the extent to which different kinds of mental states (beliefs, desires, emotions, etc.) differ in perceived controllability relative to clear-cut behavioral benchmarks. We were also interested in the extent to which attributions of control over objectionable mental states predict downstream social judgments of blame and responsibility for them.

Prior to the present inquiry, past research and theorizing on these questions yielded a state of conflict and uncertainty. With respect to attributions of control, 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., 2018). However, work in linguistics appeared to point in the opposite direction (Katz & Postal, 1964), as did empirical research within social psychology, suggesting that mental states are seen as largely uncontrollable (Gilovich & Regan, 1986; Malle & Knobe, 1997b). Indeed, some social psychologists had theorized that people perceive many mental states as things that “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.

The state of confusion is only reinforced when it comes to differences between mental states. As reviewed in the Introduction, a variety of predictions emerge from past anthropological and philosophical work on the topic. One commonality between these predictions is the idea that particular thinking processes are attributed a (relatively) high degree of controllability, and that desires, evaluations, and emotions are attributed a low degree of controllability. Finally, with respect to the relation between control and blame (or responsibility), some past research does suggest such a link between these constructs, such that greater control over mental states leads to heightened blame (Cohen & Rozin, 2001), but again, the existing investigations consider only a very limited range of stimuli. In sum, despite longstanding philosophical interest in the topic of mental control, and despite its relevance to a number of real-world psychological phenomena (see Introduction), the existing evidence is highly piecemeal, and not sufficient to substantiate any general conclusions about perceptions of mental

state control. Our goal was to address this problem through a series of studies that used triangulating methodologies.

Across six studies we confirmed four 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—emotions, desires, beliefs, evaluations, and thoughts, among others—the typical amount of control our subjects attributed was at or above the midpoint of a 7-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). The one exception to this were emotions which, in Studies 1–3, were judged as more uncontrollable and unintentional than controllable and intentional. However, 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) across all studies which compared these categories. Thus, although 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), no category of mental states was regarded as “just happening” to people.

These results call into question what we perceive to be the dominant view among scholars; namely, that mental states are judged by ordinary people to be passively experienced and basically uncontrollable (see Introduction). This view has often been accompanied by predictions that people will explain and evaluate mental states in much the same way that they explain and evaluate uncontrollable behaviors (e.g., Gilovich & Regan, 1986; Malle & Knobe, 1997b; Sabini & Silver, 1998). Our findings supplant this view and generate a novel and contrasting prediction: because others’ mental states are often seen as quite controllable, people will often explain and evaluate them similarly to the way they explain and evaluate voluntary behaviors.

Our second major finding was that there is systematic variation in perceived control across different types of mental states. Studies 1–3 showed that, when the content and context of mental states were allowed to vary so that people considered prototypical mental states from each category, the following stable ordering of control emerged: emotions were seen on average as less controlled than desires, which in turn were seen as less controlled than beliefs and evaluations, which were seen as less controlled than thoughts (e.g.,

deliberations and imaginings). Interestingly, beliefs and evaluations, which may seem like very different kinds of mental states (the latter often being grouped with other kinds of affective experiences), were reliably attributed similar degrees of control and intentionality.

It is useful to contrast these findings with previous claims about how control differs across mental states. As noted in the Introduction, scholars have variously claimed that beliefs are more controllable than other mental states (D'Andrade, 1987; Turri et al., 2018), that desires are less controllable than emotions (D'Andrade, 1987), or have otherwise failed to distinguish between emotions and evaluative attitudes (e.g., D'Andrade, 1987; A. Smith, 2005). These predictions met with varying degrees of success in our data. Beliefs were indeed judged to be one of the most controllable sorts of mental state, though they clearly do not stand alone in this regard, as data from all of our studies showed. Evaluations, imaginings, and deliberations were judged just as controllable, if not more controllable, than beliefs. Desires were judged as less controllable than beliefs, although only when their contents were allowed to vary, and not when their content was tightly controlled. Finally, emotions were not judged as equivalent to evaluations in controllability; instead, they were reliably judged to be less controllable.

However, this apparent ordering of perceived mental state control is complicated by our third finding, that much of the difference between mental state categories owes to the context and content associated with those categories. In Studies 4–6, we held the content and context surrounding each mental state constant. This was an especially conservative procedure that was designed to determine whether the ordering observed in our earlier studies arose from systematic differences in the envisaged content (and context) of the various mental states, or from conceptual differences intrinsic to each category. When we did this, the only difference from the earlier studies that remained significant was the lower ratings of intentional control for emotions. This difference occurred for judgments of intentional choice and control, though not for judgments of the ability to stop having the mental state (see Studies 5 and 6). More generally, as our meta-analysis revealed, we typically observed smaller differences between mental state categories when content and context were controlled.

Taken together, these results suggest that the lower ratings of control for emotions reflect a fundamental (or intrinsic) difference between emotions and other mental state categories (at least insofar as judgments of choice and control are concerned)—that is, something in the very nature of being an emotion generates an impression of lowered mental control as compared with the other mental states. In contrast, the differences we observed between the other mental states in the earlier studies seem instead to reflect systematic differences in the imputed content (and context) of those mental states—that is, the fact that beliefs, desires, thoughts, and evaluations typically have different kinds of everyday contents.

In our view, this does not make the differences observed in Studies 1–3 any less “real.” The conservatism of the procedure we used in Studies 4–6 has its advantages in terms of experimental control, but it also comes at the cost of sacrificing representativeness. In everyday life, there is no reason to think that desires, beliefs, and thoughts usually do have the same content—their content likely does vary in systematic ways, as better reflected by

the subject-generated materials used in Studies 1–2. In this sense, we think the results from these earlier studies do provide valid information about the *prototypical* perceived controllability of a variety of everyday mental state categories. However, we should not infer that, because prototypical control varies by mental state, this reflects intrinsic properties of the mental states in question, as once content and context are strictly controlled, many of these differences cease to exist.

### Control, Responsibility, and Character

Our fourth, and final, major finding was that there is a positive relationship between judgments of mental state control and moral judgment. As noted in the Introduction, control has proven to be an important predictor of moral judgments about bad behaviors (e.g., Alicke, 2000; Malle et al., 2014), achievements (e.g., Weiner, 1995), and mental illnesses (see Haslam & Kvaale, 2015, for a review). Yet it has remained unclear whether control plays a similar role for everyday mental states. Accordingly, we measured judgments of responsibility and blame for mental states in several of our studies, as well as measures of the extent to which mental states provide information that is diagnostic of their possessor's moral character (character-relevance). Study 1 investigated judgments of mundane mental states that were not explicitly designed to be bad or immoral. Study 2 investigated objectionable mental states of others that our subjects had encountered in their everyday lives. Study 3 investigated judgments of responsibility for highly abstract mental states. Study 5 investigated experimenter generated mental states that were designed to be antisocial, and Study 6 investigated prosocial mental states. In every one of these studies, a clear pattern emerged: ratings of intentionality and control strongly correlated with judgments of blame and responsibility, as well as with judgments of character-relevance, but the relationship was consistently stronger for blame and responsibility than it was for character-relevance.

This pattern of relationships between control, character, and blame is noteworthy for several reasons. First, it suggests that, 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. 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, Goodwin, & Goodman, 2016; Szczurek, Monin, & Gross, 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, none of these studies measured or manipulated the controllability or intentionality of the mental states in question, nor did these prior studies measure judgments of blame for those mental states. Our results suggest that doing both of these things would provide a more complete picture of people's reactions: when confronted with someone holding a deviant attitude, people may be concerned not only with whether the person is good or bad, but also with whether the person chose that attitude or can now change it, and thus whether they warrant blame for it. We conjecture that judgments of this sort are often made during interpersonal conflicts, particularly

those that involve ideological disagreement or the perception of inappropriate or unjustified attitudes.

Second, it extends prior findings by comprehensively showing that control and intentionality predict moral judgments of mental states, just as they predict moral judgments of behavior. Although this had previously been documented for a limited class of mental states (Cohen & Rozin, 2001), the current findings document this relationship over a much wider array of mental states. Moreover, whereas Cohen and Rozin (2001) focused on the relationship between mental state control and judgments of bad character, and to some extent, generic judgments of the “badness” of holding an objectionable mental state, our findings go further by showing that control also predicts the more specific moral judgment of blame. Because blame predicts emotional reactions such as anger, and behavioral reactions such as moral censure and punishment (e.g., Malle et al., 2014, and Weiner, 1995, for reviews), we hypothesize that judgments of control would predict similar behavioral reactions to mental states; this hypothesis awaits testing.

Third, the positive relationship between control and character relevance, although small, warrants attention because it conflicts with some important prior findings in related areas. 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 of the dangerousness of individuals suffering from those illnesses, 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. The reason for this discrepancy is not clear to us, but it opens up questions for future research.

### Limitations and Future Directions

The preceding descriptions of the ordering of mental state categories in terms of perceived control oversimplifies the data in a notable way; namely, it ignores the fact that there were quite sizable differences within each mental state category (see supplemental materials Figure S3–S4). The variation within each mental state category may itself spring from two different sources—the fact that each category included subtly different mental state concepts, and the fact that content also varied across the items within each category. Taking the category of belief as an example, in Studies 1 and 3, we found that beliefs that were more spontaneous or affectively tinged, such as intuitions ( $M = 4.28, SD =$

1.41), impressions ( $M = 4.11, SD = 1.64$ ), and feelings (e.g., “he felt that . . .,”  $M = 4.00, SD = 1.64$ ) tended to be rated lower in control than beliefs that were framed as judgments ( $M = 6.10, SD = 0.89$ ), acceptances ( $M = 5.65, SD = 1.14$ ), and conclusions ( $M = 5.55, SD = 1.54$ ). With regard to content, there may also be some systematic factors that gave rise to this within-category variation. For instance, there may be differences in perceived control between perceptual and abstract beliefs (Turri et al., 2018). Perceptual beliefs (e.g., “It is raining outside”) have an immediate and spontaneous quality (Ryan, 2003), which may make them seem less controllable than more abstract beliefs (e.g., “The plants will grow because of the rain”). Future work investigating judgments of control (or other judgments of agency) will likely benefit from exploring these differences.

Another limitation concerns how we measured perceived control. Although we asked about different kinds of control, including intentional control, choice, preventability, and the ability to undo a mental state, we did not assess how people think such control can be executed (see Alston, 1988, and Wegner & Pennebaker, 1993, for a discussion). There may be differences between mental states in how directly control can be executed. For instance, some mental states may be thought to be controllable directly (i.e., through an act of will), whereas others may be thought to be changeable only via the modification of other mental states, or through changes in behavior, habits, or the environment. Perceived differences in how control can be executed are likely masked by ratings of intentionality or controllability. For instance, *jumping* and *cleaning a mansion* can both be done entirely intentionally, but the former can be executed directly and immediately whereas the latter cannot. Determining how people think that changes to mental states can be brought about will provide a more complete picture of the lay understanding of mental control.

Future work should also investigate whether these results replicate outside the restricted sample of U.S. students and adults we used in our studies. Although restricted sampling is a problem plaguing many findings in psychology, there is some evidence suggesting that judgments of mental control may be especially prone to variation among different cultures (religious or otherwise; e.g., Lillard, 1997). In their seminal paper, Cohen and Rozin (2001) reported that Christian students regarded mental states such as not liking one’s parents, or thinking about having an extramarital affair as more controllable than did Jewish students, which may be related to Christians having many more religious proscriptions about mental states than do Jews. Cross-cultural investigation is therefore an important task for future research.

The present research is also limited insofar as it only investigated lay judgments of *other people’s* control over mental states; it does not investigate attributions of control for the self. In fact, in work we report elsewhere, we find that people routinely judge their own mental states to be less controllable than they judge others’ to be (e.g., people judge that others can voluntarily change their belief in God, whereas they themselves cannot; Cusimano & Goodwin, 2018). This finding is surprising when one considers that people also attribute to themselves more free will over behavior than they attribute to others (Pronin & Kugler, 2010). The reversal appears to result from a special characteristic of mental states, which is that people have unique, introspective access to the particular constraints operating on their own mental states, but often fail to impute those same internal constraints to others.

Finally, our studies do not address how lay judgments of mental state control relate to any individual's actual capacity to control what they feel, want, believe, think and so on. Some scholars have argued that beliefs are in fact not within people's control (e.g., Alston, 1988; James, 1937), which suggests that the high attributions of control in our studies are an error in lay people's judgments. In a similar vein, a great deal of work suggests that people have far less control over their thoughts than they ordinarily assume (see Wegner & Pennebaker, 1993, for a collection of important essays on this topic). Our studies are silent on the question of how much control people actually have over their own mental states. Investigating the actual control that people have over their mental states, and how this relates to people's default beliefs about controllability, remains an important avenue for future research.

### Conclusion

Judgments of control influence how we think about and react to a wide variety of social phenomena; however, little work has investigated whether people judge the contents of other people's minds as controllable. The present results show that people do tend to see most mental states as moderately controllable, believing that others choose the beliefs, desires, emotions, and attitudes they have to a considerable extent, and could opt not to have such mental states if they chose to. The degree of control varies by mental state, with emotions systematically perceived to be lower in control than other kinds of mental states, though not as low as unintentional behaviors. However, it would be an overstatement to claim that perceptions of control are firmly constrained by mental state category, as we observed variation within categories, and much variation that was explained by the mental states' content and surrounding context. Finally, judgments of control predict judgments of blame for mental states, just as they do for behaviors.

These findings set the stage for a variety of future investigations. Moving forward, researchers should investigate whether these findings generalize to other populations, what factors influence perceived mental state control both within and across mental state categories, how people think such control is executed, how and why judgments of self and other control over mental states might differ, and how ordinary people react to and sanction deviant mental states in everyday life. We have begun investigating some of these questions, but much remains to be done.

### References

- Adams, R. M. (1985). Involuntary sins. *The Philosophical Review*, *94*, 3–31. <http://dx.doi.org/10.2307/2184713>
- Akhtar, O., & Wheeler, S. C. (2016). Belief in the immutability of attitudes both increases and decreases advocacy. *Journal of Personality and Social Psychology*, *111*, 475–492. <http://dx.doi.org/10.1037/pspa0000060>
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574. <http://dx.doi.org/10.1037/0033-2909.126.4.556>
- Alston, W. P. (1988). The deontological conception of epistemic justification. *Philosophical Perspectives*, *2*, 257–299. <http://dx.doi.org/10.2307/2214077>
- 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*, 586–593. <http://dx.doi.org/10.1111/j.1467-9280.2009.02330.x>
- Arpaly, N., & Schroeder, T. (2013). *In praise of desire*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199348169.001.0001>
- 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>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. West Sussex, UK: Wiley.
- Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*, 237–273. [http://dx.doi.org/10.1016/0010-0277\(83\)90006-9](http://dx.doi.org/10.1016/0010-0277(83)90006-9)
- Cohen, A. B., & Rankin, A. (2004). Religion and the morality of positive mentality. *Basic and Applied Social Psychology*, *26*, 45–57. [http://dx.doi.org/10.1207/s15324834basp2601\\_5](http://dx.doi.org/10.1207/s15324834basp2601_5)
- Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology*, *81*, 697–710. <http://dx.doi.org/10.1037/0022-3514.81.4.697>
- Cusimano, C., & Goodwin, G. P. (2017a). *Bad attitudes are treated like bad acts in everyday moral evaluation*. Unpublished manuscript.
- Cusimano, C., & Goodwin, G. P. (2017b). *Folk attributions of control and intentionality over mental states*. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Cusimano, C., & Goodwin, G. P. (2018). *The believer-observer asymmetry: People judge others to have more control over their beliefs than they, themselves do*. Unpublished manuscript.
- 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>
- 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*, 497–505. <http://dx.doi.org/10.1080/01973533.2013.840632>
- Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry*, *6*, 267–285. [http://dx.doi.org/10.1207/s15327965pli0604\\_1](http://dx.doi.org/10.1207/s15327965pli0604_1)
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273. <http://dx.doi.org/10.1037/0033-295X.95.2.256>
- Eid, M., & Diener, E. (2001). Norms for experiencing emotions in different cultures: Inter- and intranational differences. *Journal of Personality and Social Psychology*, *81*, 869–885. <http://dx.doi.org/10.1037/0022-3514.81.5.869>
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *The Journal of Economic Perspectives*, *30*, 133–140. <http://dx.doi.org/10.1257/jep.30.3.133>
- 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*, *147*, 1170–1190. <http://dx.doi.org/10.1037/xge0000396>
- Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, *60*, 328–331. <http://dx.doi.org/10.1198/000313006X152649>
- 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>
- 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*, 42, 1077–1091. <http://dx.doi.org/10.1177/0146167216651408>
- Haider-Markel, D. P., & Joslyn, M. R. (2008). Beliefs about the origins of homosexuality and support for gay rights: An empirical test of attribution 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, 1–36. <http://dx.doi.org/10.1111/j.1559-1816.2003.tb02071.x>
- Hartshorne, J. K., O'Donnell, T. J., Sudo, Y., Uruwashi, M., Lee, M., & Snedeker, J. (2016). Psych verbs, the linking problem, and the acquisition of language. *Cognition*, 157, 268–288. <http://dx.doi.org/10.1016/j.cognition.2016.08.008>
- Haslam, N., & Kvaale, E. P. (2015). Biogenetic explanations of mental disorder the mixed-blessings model. *Current Directions in Psychological Science*, 24, 399–404. <http://dx.doi.org/10.1177/0963721415588082>
- Heider, F. (1958). *The psychology of interpersonal relations* (Vol. IX). Hoboken, NJ: Wiley. <http://dx.doi.org/10.1037/10628-000>
- Huddleston, R. (1970). Some remarks on case-grammar. *Linguistic Inquiry*, 1, 501–511.
- James, W. (1937). The will to believe. In *The will to believe, and other essays in popular philosophy* (pp. 1–31). London, UK: 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>
- Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, 3, 81–123. <http://dx.doi.org/10.1080/02699938908408075>
- Kappes, A., & Schikowski, A. (2013). Implicit theories of emotion shape regulation of negative affect. *Cognition and Emotion*, 27, 952–960. <http://dx.doi.org/10.1080/02699931.2012.753415>
- Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.
- Kneeland, E. T., Nolen-Hoeksema, S., Dovidio, J. F., & Gruber, J. (2016). Emotion malleability beliefs influence the spontaneous regulation of social anxiety. *Cognitive Therapy and Research*, 40, 496–509. <http://dx.doi.org/10.1007/s10608-016-9765-1>
- Kvaale, E. P., Gottdiener, W. H., & Haslam, N. (2013). Biogenetic explanations 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>
- 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, 1225–1237. <http://dx.doi.org/10.1037/0022-3514.74.5.1225>
- 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>
- Lillard, A. S. (1997). Other folks' theories of mind and behavior. *Psychological Science*, 8, 268–274. <http://dx.doi.org/10.1111/j.1467-9280.1997.tb00437.x>
- Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.
- Malle, B. F. (2014). *F. Ex: A coding scheme for folk explanations of behavior*. Unpublished manuscript, Brown University, Providence, RI.
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, 25, 147–186. <http://dx.doi.org/10.1080/1047840X.2014.877340>
- Malle, B. F., & Knobe, J. (1997a). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121. <http://dx.doi.org/10.1006/jesp.1996.1314>
- Malle, B. F., & Knobe, J. (1997b). Which behaviors do people explain? A basic actor-observer asymmetry. *Journal of Personality and Social Psychology*, 72, 288–304. <http://dx.doi.org/10.1037/0022-3514.72.2.288>
- Malle, B. F., & Knobe, J. (2001). The distinction between desire and intention: A folk-conceptual analysis. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 45–67). Cambridge, MA: MIT Press.
- Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Belknap Press. <http://dx.doi.org/10.4159/harvard.9780674421288>
- 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>
- Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, 1, 29–50. <http://dx.doi.org/10.1080/02699938708408362>
- 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, 795–809. <http://dx.doi.org/10.1037/a0012710>
- Pronin, E., & Kugler, M. B. (2010). People believe they have more free will than others. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 22469–22474. <http://dx.doi.org/10.1073/pnas.1012046108>
- Pylkkänen, L. (1999). On stativity and causation. In C. Tenny & J. Pustejovsky (Eds.), *Events and grammatical objects* (pp. 417–445). Stanford, CA: CSLI Publications.
- 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>
- Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, 121, 192–218. <http://dx.doi.org/10.1037/0033-2909.121.2.192>
- Russell, J. A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology*, 45, 1281–1288. <http://dx.doi.org/10.1037/0022-3514.45.6.1281>
- Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies*, 114, 47–79. <http://dx.doi.org/10.1023/A:1024409201289>
- Sabini, J., & Silver, M. (1998). *Emotion, character, and responsibility*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195121674.001.0001>
- Sankowski, E. (1977). Responsibility of persons for their emotions. *Canadian Journal of Philosophy*, 7, 829–840. <http://dx.doi.org/10.1080/00455091.1977.10716199>
- 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](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 students. *Cognitive Therapy and Research*, 39, 120–139. <http://dx.doi.org/10.1007/s10608-014-9652-6>
- Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology*, 71, 549–570. <http://dx.doi.org/10.1037/0022-3514.71.3.549>

- Smith, A. M. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics, 115*, 236–271. <http://dx.doi.org/10.1086/426957>
- Smith, A. M. (2008). Control, responsibility, and moral assessment. *Philosophical Studies, 138*, 367–392. <http://dx.doi.org/10.1007/s11098-006-9048-x>
- Smith, H. M. (2011). Non-tracing cases of culpable ignorance. *Criminal Law and Philosophy, 5*, 115–146. <http://dx.doi.org/10.1007/s11572-011-9113-1>
- 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>
- Szczurek, L., Monin, B., & Gross, J. J. (2012). The stranger effect: The rejection of affective deviants. *Psychological Science, 23*, 1105–1111. <http://dx.doi.org/10.1177/0956797612445314>
- Tetlock, P. E., Kristel, O. V., Elson, S. B., 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*, 853–870. <http://dx.doi.org/10.1037/0022-3514.78.5.853>
- Turri, J., Rose, D., & Buckwalter, W. (2018). Choosing and refusing: Doxastic voluntarism and folk psychology. *Philosophical Studies, 175*, 2507–2537.
- Wegner, D. M., & Pennebaker, J. W. (1993). *Handbook of mental control*. Englewood Cliffs, NJ: Prentice Hall.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York, NY: Guilford Press.
- Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

## Appendix A

### Stimuli Used in Study 1

---

He dropped his cup on the floor  
 She dropped her phone on the ground  
 She dropped she dropped her car keys  
 He dropped the bowling ball  
 He dropped coffee on his white dress shirt

He fell off of the swing  
 He fell off of the horse  
 She fell off of her bicycle  
 He fell off of the ladder  
 She fell off of the monkey bars

He ran into a wall  
 He ran into his neighbor's mailbox  
 She ran into the glass door  
 He ran into the bicyclist  
 He ran into a friend outside of the building

He slipped on a patch of ice  
 He slipped on the damp concrete  
 She slipped on a banana peel  
 She slipped on the black ice  
 He slipped on the wet floor

He tripped over his own feet  
 He tripped over the tree branch  
 He tripped over a rock on the path  
 She tripped over a cat  
 He tripped over the cord on the floor

She believed that she did well on the exam  
 He believed that God exists  
 She believed that unicorns were real  
 She believed that her grandpa could make a comeback after his stroke  
 He believed that the girl liked him

---



---

She felt that she could do better  
 She felt that she was one of the only people with depression  
 She felt that she had done well on the exam  
 He felt that life just wasn't fair  
 He felt that things would work out

He thought that everything would work out fine  
 She thought that she wasn't good enough for him  
 He thought that life is hard  
 He thought that she was pretty  
 He thought that the death penalty was unjust

She concluded that her hypothesis was correct  
 He concluded that she must like him  
 She concluded that her brother was no longer coming  
 She concluded that her friend was a psychopath  
 He concluded that it was time to leave

She considered going to the dance  
 She considered becoming an accountant  
 She considered buying some ice cream  
 He considered dropping out of college  
 He considered skipping class to go to a concert

She deliberated about what she would say  
 He deliberated about what he wanted to do for his 21st birthday  
 She deliberated about which classes she should take  
 She deliberated about asking him to the formal  
 She deliberated about going to the party versus studying

He speculated about where he would be in three years  
 He speculated about why his teacher cancelled class  
 She speculated about what her friends were doing right now  
 He speculated about the ending of the mystery novel  
 She speculated about how her friend really felt

---

(Appendices continue)

---

She thought about what had happened to her that day  
 She thought about eating another piece of cake  
 He thought about how much he loved her  
 He thought about what the future might hold for him  
 He thought about his family

He craved chocolate  
 She craved approval  
 She craved mom's cooking  
 She craved pizza  
 She craved for Indian food

She desired a chocolate chip cookie  
 He desired her  
 She desired to meet her favorite singer  
 He desired to be outside rather than in class  
 She desired to take a vacation

She hoped for a good figure  
 She hoped for a delicious dinner  
 He hoped for another chance to make things right with her  
 He hoped for a good result on his examination  
 She hoped for a good life

He wanted to go home  
 She wanted a new car for her birthday  
 He wanted to go to the concert  
 He wanted to tell them how he was feeling  
 He wanted an ice cream cone

She felt angry with her boyfriend  
 He felt angry because he failed the exam  
 He felt angry when they called him the wrong name  
 She felt angry that he lied to her  
 He felt angry when he thought about how racism still exists in this country

He felt anxious about his test grade  
 He felt anxious about performing his music  
 She felt anxious about the upcoming interview  
 She felt anxious about her upcoming exam  
 He felt anxious about telling the truth

She felt embarrassed about showing up late to class  
 She felt embarrassed when she tripped on the sidewalk  
 She felt embarrassed when the girl made fun of her  
 She felt embarrassed by her mistake  
 He felt embarrassed when he couldn't answer the question

She felt happy because it was her birthday  
 She felt happy when she was with her family  
 He felt happy that his deal went through  
 She felt happy about her grade  
 He felt happy when the pizza arrived

She hated the smell of cut grass  
 She hated the girl with the brown hair  
 She hated her boyfriend  
 He hated his stat class  
 He hated drinking tea

---

He loved his mom  
 She loved God  
 He loved chocolate  
 He loved his wife  
 He loved to sing

She valued money and power  
 She valued hard work and respect  
 He valued her opinion  
 She valued her friendship with him  
 He valued honesty

He appreciated her help  
 She appreciated the flowers he bought her  
 He appreciated his friend's sincerity  
 She appreciated her mother  
 He appreciated the professor

She imagined getting an A on her test  
 She imagined that she could be an astronaut  
 He imagined going to the theatre with the cute girl in his class  
 He imagined what that street looked like in 1817  
 She imagined her future house

He visualized his new home in California  
 She visualized their future  
 He visualized how he was going to be on a beach soon  
 She visualized the picture in her head  
 She visualized what it would be like to fly

He intended to lose 10 pounds by Christmas  
 She intended to break up with him  
 She intended to finish her asparagus  
 She intended to sleep early  
 He intended to finish his homework tonight

He planned to go to the gym tomorrow morning  
 She planned to write a book  
 She planned to go to the mall on Friday  
 She planned to take a shower after work  
 She planned to do better on the next test

She resolved to eat salad tomorrow  
 He resolved to see the Pope  
 She resolved to make more friends this year  
 She resolved to finish the essay  
 She resolved to be a better person

He had the goal to get a promotion at work  
 She had the goal to win the championship game  
 He had the goal to receive a perfect score on the exam  
 She had the goal to become a doctor  
 He had the goal to help others

He ate his dinner  
 She ate only dinner on Monday  
 He ate the sweet green salad  
 She ate another piece of cake  
 He ate oatmeal for breakfast every day

---

*(Appendices continue)*

---

He avoided her gaze  
 She avoided the dark alley  
 He avoided scheduling an appointment with the dentist  
 She avoided telling her mentor about her struggles  
 She avoided her ex

He played with the toys  
 She played with her dog  
 She played with her cat  
 He played with his little cousins  
 He played with his younger brother on the playground

He said hello  
 He said that he would never leave her  
 She said what other people wanted to hear  
 She said that the trip was fun  
 He said he was sorry

He searched for the best deal online  
 He searched for his wallet  
 She searched for her keys  
 She searched for a job position  
 He searched for the homework answers on google

He forgot how beautiful life could be  
 She forgot that she had scheduled a meeting during lunchtime  
 He forgot if he wore socks  
 She forgot his name  
 She forgot about her assignment due at 10

He remembered when he was a child  
 She remembered the time she fell down the stairs  
 He remembered everything that he had with her

---



---

He remembered that he left the door unlocked  
 He remembered his mom's homemade jelly

He fainted at the sight of blood  
 He fainted of exhaustion  
 She fainted from dehydration  
 She fainted when her boyfriend proposed to her  
 He fainted in the middle of the courtyard

She shivered because it was cold  
 He shivered because he had a very bad cold  
 She shivered at the thought of him  
 He shivered at the thought of going outside

She shivered when the wind blew

He sneezed after smelling the flower  
 She sneezed when she entered the dusty old room  
 He sneezed after playing with the dog  
 He sneezed in the middle of class  
 He sneezed as he was driving

She sweated through her cotton t-shirt  
 She sweated in the 99-degree heat  
 He sweated during the game  
 He sweated when he played soccer  
 She sweated when she went on a run

She yawned during the lecture  
 He yawned when he was tired  
 She yawned before falling asleep  
 He yawned after he woke up from a nap  
 She yawned at the book

---

*(Appendices continue)*

**Appendix B**  
**Sample Responses From Study 2**

State	Sample scenario produced by subject
Emotion	<p>He seemed irritable and upset that I wanted to discuss future plans that we can't be certain about.</p> <p>He was very angry and irritated for no apparent reason.</p> <p>The person wasn't displaying any emotion at the death of a relative.</p> <p>The individual was exhausted, irritated and feeling upset.</p> <p>He simply was apathetic to the passing of a childhood pet.</p> <p>She was feeling depressed and having suicidal thoughts.</p> <p>The person's emotion was angry, frustrated, outraged.</p>
Desire	<p>This person wanted to cheat on his significant other.</p> <p>The person wanted to do dangerous activities.</p> <p>She was being selfish and wanted to be the center of attention in a reunion made for another friend.</p> <p>Wanted me to have a bad relationship with my father.</p> <p>To cheat on a final exam.</p> <p>She had the desire to cut herself.</p>
Evaluation	<p>He expressed his dislike for gay couples showing affection in public.</p> <p>He had a negative perception of immigrants in the community.</p> <p>Lack of gratitude.</p> <p>He seemed to enjoy the comments his coach made to the girls on the team about their weight/eating habits.</p> <p>He had a pro-Trump attitude.</p>
Belief	<p>He didn't care about the well-being of someone else.</p> <p>The person felt as if the way a woman dresses allows people to judge her or [sic] for certain violent behaviors to happen to her.</p> <p>That doing drugs is okay.</p> <p>He was expressing the belief that he thinks driving under the influence of marijuana is ok.</p> <p>She strongly believed that depression was not a mental issue and could be overcome by simply "manning up."</p> <p>He believed Donald Trump would be the best president for the U.S. despite his various indiscretions.</p> <p>The person believed that gay marriage was wrong.</p>

*(Appendices continue)*

### Appendix C

#### Mental States Used in Study 3

---

Behavior foils 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  Mental states 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
--	---

---

*(Appendices continue)*

## Appendix D

### Stimuli Used in Study 5

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 was not 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 was not in even more pain.

*(Appendices continue)*

## Appendix E

### Stimuli Used in Study 6

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, she 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 should survive.

Upon hearing this, Paul . . .

- feels happy/excited that his mother should survive.
- wants/desires his mother to survive.
- thinks/believes that life would be better if his mother survived.
- likes/loves the fact that his mother should 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 but has called to tell her parents she and her long-time boyfriend just got engaged.

When he hears this, James . . .

- feels happy/excited that his daughter is getting married.
- wants/desires his daughter to get married.
- thinks/believes that it is good that his daughter is getting married.
- likes/loves the fact that his daughter is getting married.

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 has captured a city and is likely going to publicly execute dozens of journalists currently hiding out there. In re-

sponse, the United Nations (UN) has launched a counter attack which, according to analysts, is likely to succeed given the UN force's superiority.

While watching all of this on the news, Wesley . . .

- feels happy/excited that that the UN's counter attack will likely succeed.
- wants/desires the UN counter attack to succeed.
- thinks/believes that it would be better if the UN counter attack succeeds.
- likes/loves the fact that the UN counter attack will likely succeed.

Amy is a college student watching a documentary about polar bears. The documentary is about whether polar bears are adapting to climate change in the Arctic, and it tracks a specific polar bear and her cubs. For most of the movie, it is not clear whether the family of bears will survive the summer. At the end, however, it is revealed that the bear family discovered an area where the ice is still thick enough for them to hunt on. The narrator indicates that, thanks to this, the bears have a better chance of making it through the summer.

As the credits role, Amy . . .

- feels happy/excited that the bears may survive the summer.
- wants/desires the bears to survive the summer.
- liked/loved the fact that the bears may survive the summer.
- thinks/believes that it is good that the bears may survive the summer.

Received May 18, 2018

Revision received October 30, 2018

Accepted November 1, 2018 ■