Mental states and control-based theories of responsibility.

Corey Cusimano[1] and Geoffrey P. Goodwin[2]

[1]Princeton University
[2]University of Pennsylvania

[Forthcoming in Thomas Nadelhoffer and Andrew Monroe (eds)., *Advances in experimental philosophy of free will and responsibility.* Bloomsbury.]

On October 13, 1988, at the start of the second presidential debate, Democratic Nominee Michael Dukakis took a question from the moderator challenging his long-held and widely-known anti-death penalty stance, "Governor, if Kitty Dukakis were raped and murdered, would you favor an irrevocable death penalty for the killer?" Dukakis, true to his principles, responded, "I think you know that I've opposed the death penalty all of my life." As he continued laying out the reasons behind his opposition, he remained practiced, professional, and composed – which was exactly the problem. Potential voters who were watching were outraged at his unemotional reaction to the thought of his wife's hypothetical rape and murder. Dukakis' popularity dropped overnight. Outrage at Dukakis did not reflect objections to his principles or even doubts about his perceived commitment to his wife. In the words of a columnist at the time, "[it] is well-known that the Governor feels about Kitty the way Antony felt about Cleopatra and Romeo about Juliet" (McGrory, 1988, November 10). Everyone knew that he loved his wife, that the question itself was incendiary, and that his answer was nonetheless a good one, supported by strong evidence and reflecting deeply-held moral commitments. Rather, the outrage reflected his "inadmissibly impersonal" reaction that failed to "merchandise his emotions" (McGrory, 1988, November 10). In short, he was blamed for having the wrong emotional reaction. This extraordinary moment in U.S. politics betrays something quite mundane: We sometimes blame others for having the wrong attitudes, beliefs, wants, and feelings.

The case of Michael Dukakis raises both a *psychological* and a *philosophical* question about the nature of blame. The psychological question asks: In everyday life, what are the intuitive preconditions for blame? In other words, what occurred in minds of the American populace that led them to think Michael Dukakis should be blamed for his emotionless reaction? The philosophical question asks: What is the most defensible theory about whether, and on what basis, someone can be said to deserve blame? In other words, assuming that Michael Dukakis's emotional reaction was bad, were people right to blame him for it? These two questions are distinct – after all, in everyday life, people may blame others in ways that seem mistaken after careful philosophical reflection. But there is also an important way in which the two questions are related, which is that philosophers have often treated everyday practices of blame as providing evidence for different philosophical positions (Knobe & Doris, 2010). Following Strawson (1962), many philosophers have constructed normative accounts of blameworthiness that aim to make the best sense of ordinary practices of blaming. Accordingly, the intuitive basis on which people blamed Dukakis might be treated as a starting point for a normative view of blame that ultimately serves to validate people's reactions.

Our focus in this chapter is why people sometimes assign blame for mental states such as Dukakis's. For reasons that we outline below, the question of mental state blame occupies a privileged position in both empirical and normative debates about moral responsibility. The reason is that while different theories of moral responsibility converge on the question of how people assign blame for behaviors, they make divergent predictions regarding whether, and on what basis, people blame others for their mental states. Thus, examining blame for mental states offers a means of adjudicating between these rival theories. Looking ahead, many theories suggest that our everyday practices of blame presuppose that people have *control* over whatever it is they are being blamed for. Yet, many philosophers point to everyday examples of mental state blame, such as people's reaction to Dukakis, as posing a counterexample to these theories. They claim that emotions (and other mental states) are not controllable, and so the fact that people regularly blame others for them suggests that normative accounts of blame should not presuppose control either. Here we offer a novel resolution to this debate. Based on recent empirical work, we conclude that scholars have mischaracterized the everyday practices of blame by assuming an incorrect basis for why people assign blame for mental states. We argue that people do readily blame others for their mental states, but they do so precisely because they think mental states are under people's control.

Control as a precondition for moral responsibility

When those around us cause harm, or act inappropriately, we tend to hold them *morally responsible* for doing so. Holding someone morally responsible, accountable, or in other words, blaming them, often involves expressing anger toward them, criticizing them, demanding that they explain themselves or make amends, or otherwise making them feel bad (Coates & Tognazzini, 2013; Malle, Guglielmo, & Monroe, 2014). These behaviors communicate to the blamed party that what they did was unacceptable, they establish our expectations moving

forward, and they provide motivation to the transgressor to fulfill those expectations. However, it is only acceptable to blame someone, and so confront them about their behavior and make them feel bad for it, if they deserve it – that is, if they are blame*worthy*. Indeed, blaming someone who does not meet this standard is itself blameworthy. This observation forms the grounds for the psychological and philosophical questions above: In everyday practice, what do people think is required for an agent to deserve blame? And, normatively, what is the most defensible, principled view of what would make someone legitimately blameworthy?

The dominant view in both psychology and philosophy appears to be that someone is morally responsible for their conduct, and so only blameworthy for bad conduct, if they had control over it. The commonly cited intuition behind this proposal is that it seems wrong to criticize someone for something that they could not have done anything to avoid. That is, it is only fair to blame people for their behavior if they could have chosen to behave otherwise; and likewise, it is only fair to blame people for unfortunate events or harms if those individuals could have stopped or prevented them from occurring. Thus, assessing others' blameworthiness is principally a matter of assessing others' choices (c.f. Fischer & Ravizza, 1998; Nelkin, 2011; Wallace, 1994; Wolf, 1990). If someone could have acted in line with good moral reasons, but chose not to do so, then they are blameworthy. And accordingly, people are blameworthy for causing bad events in the world insofar as those events reflect morally bad choices that they made.

A great deal of evidence from the past 50 years attests to the role of control in everyday evaluations of blameworthiness (see Alicke, 2000; Malle et al., 2014; Weiner, 1995). Early empirical studies of blame measured perceived culpability for harm based on the various ways that a harmful outcome could be traced back to someone's decision making (Fincham & Jaspers, 1980; Weiner, 1995). The results of these studies provided strong support for control-based views. People were shown to treat intentional harms as highly blameworthy, with blame decreasing as an agent's control over the harm decreased, and as it reflected their decision making less and less. For instance, if someone did not intentionally bring about harm, but behaved with disregard for harm they knew was likely, they received lower (but still substantive) blame than someone who caused harm intentionally. And if someone caused some harm that was unforeseeable, then they tended to receive no blame at all (Fincham & Jaspers, 1979; Shultz, Schleifer, & Altman, 1981; Shultz, Wright, & Schleifer, 1986).

Later studies showed that people actively seek out information about control when evaluating someone's blameworthiness (Alicke, 2000; Malle, et al., 2014). For instance, Guglielmo and Malle (2017) found that when an observer learns that another agent caused some harm, they seek out information about whether the agent chose to do so intentionally. And, after learning that person did not intentionally cause harm, observers then tend to seek out information about the person's past (controllable) choices, and then evaluate blameworthiness based on whether that person could have prevented the harm. Furthermore, relative to the blame people assign when the intentionality of a norm violation is ambiguous, people increase blame when they learn that that the violation was intentional (and therefore more controllable) and decrease

blame when they subsequently learn that it was unintentional (and therefore less controllable; Monroe & Malle, 2019).

Finally, psychologists have routinely found that people defend themselves from blame by preferentially citing factors relating to their personal control over their behavior or the outcome in question. When held accountable for some behavior, people will argue that their behavior was unintentional, that they could not have foreseen the consequences, or that they could not have behaved otherwise, in order to convince others they are not blameworthy (Markman & Tetlock, 2000; Weiner, Amirkhan, Folkes, & Verette, 1987). Indeed, even when people know that they had control over their behavior or the outcome in question, they withhold that information from others and (deceptively) argue that they lacked control in order to avoid blame (Weiner, Figueroa-Munoiz, & Kakihara, 1991). And recently, McNeer and Machery (2019) found that most people (around 80%) explicitly endorse control as a necessary precondition of blameworthiness when asked about it in the abstract. Thus, not only do people's intuitions about blame reflect considerations of control, but people are explicitly aware of the importance of control and use it as a basis for assigning others blame and negotiating their own liability.

As noted above, these findings about the everyday practice of blame inform normative theories. For instance, Wallace (1994), who offers one of the definitive accounts of control-based theories of moral responsibility, motivates his theory by considering "our ordinary judgments of excuse and exemption from responsibility" (p. 15). The resulting moral theory then tries to honor and explain everyday individuals' adherence to control as a precondition of blame. Of course, this is not to say that normative theories do (or should) reflect a simple polling of everyday intuitions. Nevertheless, many normative moral theories get off the ground by trying to identify and make the most sense out of the implicit commitments that characterize everyday practice. The empirical data presented above strongly suggest that control comprises such a commitment, and that normative theories that want to capture everyday practices of blame need to incorporate control in some manner.

Nevertheless, both psychologists and philosophers have challenged the dominance of control in everyday moral judgment. The social psychology literature has focused on a few narrow cases, such as blame for severe accidental harms.[1] However, the most potent challenge against control, to which we now turn, comes from moral philosophy.

The mental state challenge

[1] Some scholars have argued that people blame others for uncontrollable harms that those individuals have caused – a phenomenon known as moral luck or outcome bias – and that this tendency constitutes a counterexample to control-based theories (Mazzacco, Alicke, & Davis, 2004; Robbenholt, 2000; Walster, 1966). We do not have space to discuss this challenge but will note that the empirical case for outcome bias is contested, with some evidence suggesting that it might result from biased or erroneous attributions of control (Kneer & Machery, 2019; Malle et al., 2014; Royzman & Kumar, 2004). Outcome bias is not a clear counterexample to control theories and so the best case against these theories reflects the topic of the current essay – blame for mental states.

The most potent challenge against control-based theories of blame reflects a common observation within moral philosophy that people appear to be blameworthy for their objectionable mental states. For instance, in one of the first and most forceful versions of this idea, Adams (1985) suggests that we commonly blame others for, "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). However, mental states of this sort seem uncontrollable – in Adam's words, *involuntary sins*. Since Adams, others have made similar observations, with each one of them contributing new examples wherein people seem to blame others for seemingly involuntary emotions, beliefs, motives, or evaluative judgments (see, e.g., Hieronymi, 2008; Pizarro, Tannenbaum, & Uhlmann, 2012; Sher, 2006; Smith, A., 2005, 2008; Smith, H., 2011; Sripada, 2017). Summarizing this challenge, which we call the "mental state challenge," Angela Smith (2008) proposes that, "in our day-to-day lives we simply take for granted that people are responsible and answerable for much more than what they voluntarily choose to do" (p. 382).

It is helpful to consider why these authors suggest mental states are worthy of blame despite their apparent uncontrollability. Notwithstanding subtle differences between them, these authors commonly appeal to an intuition that at least some mental states indicate something *morally significant* about a person. For instance, someone's attitudes might tell us whether they have good or faulty moral character (e.g., Pizarro & Tannenbaum, 2012), or might reveal something about their "moral personality," which might comprise their commitments, cares, or the way they assess moral reasons (e.g., Hieronymi, 2008; Smith, A., 2005; Smith, H., 2011). Accordingly, when people judge whether someone is blameworthy, they do so in part by judging whether some objectionable, morally significant part of that person has been revealed. Moreover, there can also be ways of learning something morally significant about someone without observing their voluntary, intentional behavior at all – such as when you learn about their emotions, beliefs, motives and other attitudes. People may be held blameworthy for objectionable mental states of this sort, not because they had control over them, but rather, because these mental states reveal morally significant character information about them.

However, the appeal of this argument is complicated by the observation that not all scholars accept the claim that people regularly blame others for their bad mental states. Indeed, other scholars have suggested that it is precisely because mental states are uncontrollable that people do not really blame others for them (Levy, 2005; Malle et al., 2014; Sabini & Silver, 1998; Sankowski, 1977; Wallace, 1994). Wallace (1994), for instance, writes that, even though people care about what desires and emotions others have just as they do what intentions others have, "intention appears to have a significance for questions of moral responsibility and blame that emotion and desire do not" (p. 124). Cushman (2015) suggests that, whereas people will punish others for causing harm that they could have prevented, people do not punish someone merely for holding an immoral desire. Sabini and Silver (1998) similarly predict that, in everyday life, people do not blame others for things that they cannot control, and that, "emotions, desires, passions, and impulses [are] beyond the will, without control" (p. 15). Instead, defenders

of control theories suppose that when people do blame others for mental states, this blame reflects their judgment that those individuals could have prevented themselves from adopting the attitude in the first place by making better prior choices (see, e.g., Malle et al., 2014; Rosen, 2004; Sankowski, 1977; for explicit discussion along these lines). So, these authors concede, people *occasionally* blame others for objectionable mental states, but only when it is clear that, had they made better prior choices, those mental states would never have arisen. Nevertheless, the more typical pattern according to proponents of control-based theories, is that mental states are beyond people's ability to control, and on that basis, that people rarely hold others seriously responsible for them. Thus, when faced with the question of mental state responsibility, these authors deny the intuition appealed to in the preceding paragraph, and instead reaffirm control-based theories: Believing that someone has poor or faulty beliefs, desires, emotions, or character, is not sufficient for holding them blameworthy for those mental states – instead, the attribution of control is a necessary prerequisite (see discussion in Levy, 2005; Sankowski, 1977).

Thus, the extant literature yields an impasse which we aim to resolve in the remainder of this essay. First, we will motivate the claim that everyday life is rife with mental state blame. The upshot will be that control theorists have been wrong to think that people do not (often) blame others for holding objectionable mental states. However, we will then review recent work showing that the everyday practice of mental state blame is best explained by people attributing substantive voluntary control to others over these states. The upshot here will be that scholars like Smith and Adams are wrong to think that people blame others for their mental states without appeal to control. We conclude that mental state responsibility is not a counterexample to control-based theories of blame, but instead, another vindication of this view.

Everyday blame for mental states

Philosophers who have argued that mental state blame poses a challenge to control-based theories have typically cited one of three mundane situations in which people blame others: (i) for having immoral or hurtful emotions, motives, or evaluative attitudes, (ii) for having objectionable beliefs, or (iii) for acting in ignorance that can be traced back to bad attitudes. Historically, psychological investigations of reactions to others' mental states in these scenarios has focused on people's tendency to avoid those who have objectionable mental states (e.g., Ames & Johar, 2009; Haidt, Rosenberg, & Hom, 2003; Skitka, Bauman, & Sargis, 2005; Szczurek, Monin, & Gross, Gross, 2012). In recent years, however, some empirical work comes closer to investigating whether people blame others for their mental states. We will briefly review this work below.

Bad attitudes and bad affective states. Philosophers who challenge control-based theories often cite blame-like reactions to others' objectionable motives and evaluative attitudes (such as liking loving, hating, disrespecting, and so on). For instance, Adams cites the example of someone who volunteers to help others but is self-righteous about doing so. This person's "sin" is, "not in what he voluntarily chooses to do," but rather, "in the motivation and attitude with which he usually does what he ought to do" (p. 5). Similarly, Sripada (2017) cites the

example of a father feeling jealous of his son's accomplishments as being blameworthy even though, "the father doesn't control his feelings of jealousy in any obvious sense" (p. 802). And Pizarro et al. (2012) suggest that a spouse who gets drunk and confesses to her husband, "that she never loved him, that he's a wimp and a pushover, and that he could never make her happy" would be blameworthy because of what her conduct reveals about her attitudes towards him even though her possessing (and confessing) these attitudes does not "meet the criteria for agency" (p. 186).

Although these cases have not yet been examined empirically, some recent empirical work does suggest people will blame others for possessing objectionable affective reactions. Gromet, Goodwin, and Goodman (2016) studied how people react to others who experience *schadenfreude*. In one study, participants read about a man who, in a fit of passion, murdered his spouse and the man with whom she was having an affair. Gromet et al manipulated whether the murderer took pleasure in the act (evidenced by his then mutilating the bodies) or not (evidenced by his then immediately driving away from the crime scene). When the murderer enjoyed the murder, participants rated him to be morally worse (and were more likely to say he was "evil"). Moreover, they were more likely recommend severe legal punishment, like the death penalty, when determining his sentencing. If we consider punishment as diagnostic of underlying judgments of blameworthiness, then it seems likely that participants in these studies also blamed the murderer for his emotional reaction to the crime. Moreover, in a separate study, mere observers who had not participated in or aided a harmful act were also frequently condemned as evil if they experienced pleasure upon learning about the harm, especially if the harm did not convey any instrumental benefit to them. Thus, in this case, they were condemned solely on the basis of an objectionable mental state. However, since condemnation is not synonymous with blame, this latter result does not directly establish that people are blamed for their objectionable mental states.

Recently, however, we directly studied people's attributions of blame towards of others who possess immoral mental states (Cusimano & Goodwin, 2019). In one study, participants read stories in which a protagonist formed an immoral emotion, desire, belief, or attitude towards something they watched or learned about. For instance, a student watched a recording of a reporter being tortured by state police, a father learned that his daughter is about to enter an interracial marriage, a son learned that his mother is in the hospital after a car accident, and in the last scenario, a man learned on TV that United Nations has staged a mission to rescue hostages from a terrorist organization. Each of these characters had immoral reactions to what they witnessed: The student liked the torture and wanted it to continue, the father objected to his daughter's interracial marriage, the son wanted his mother not to survive the car accident, and the man wanted the rescue mission to fail. Although we varied what specific kind of mental state the target character possessed (e.g., an immoral feeling, desire, belief, or other attitude), this variation had little effect on attributions of blame (Figure 1). Across all objectionable attitudes, and across all scenarios, participants judged that the target character's reaction was immoral, and that they were blameworthy for having it.

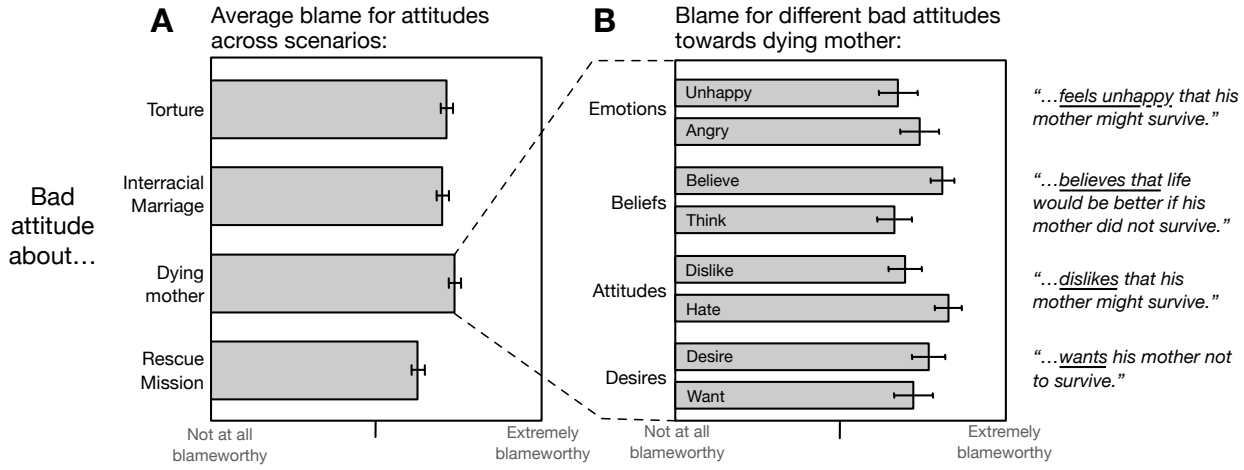

Figure 1. Mean attributions of blameworthiness for immoral mental states (Cusimano & Goodwin, 2019, Study 4; error bars represent standard errors). (A) Attributions of blameworthiness across four kinds of immoral attitudes, averaged over all mental states. (B) Attributions of blameworthiness in the "dying mother" scenario separated by mental state type. Examples of bad attitudes are in the right-most column.

Immoral, irreligious, and illogical beliefs. Many philosophers and psychologists believe that, by and large, beliefs are uncontrollable (c.f. Alston, 1988; Epley & Gilovich, 2016; see discussion in Cusimano & Goodwin, 2020, and Turri, Rose, & Buckwalter, 2018). On this basis, some have argued that it is inappropriate to demand that people hold certain beliefs or to blame them for failing to hold the right beliefs (Alston, 1988). However, others have argued that the fact that people do seem to hold others responsible for bad beliefs constitutes yet another example of how, in everyday life, people hold others responsible for things they cannot control (Adams, 1985; Hieronymi, 2008).

The current state of ideological conflict in contemporary society suggests that Adams and Hieronymi are correct that people blame others for beliefs they perceive to be objectionable. For instance, people are openly prejudiced against others who hold opposing moral and political beliefs (Brandt et al., 2014; Crawford et al., 2017; Haidt et al., 2003). And indeed, sometimes this prejudice takes the form of blame, such that people openly criticize and feel angry at others who hold opposing ideological attitudes (Garret & Bankert, 2018; Ryan, 2014).

Both experimental evidence and historical observations also suggest that people criticize, feel disdain towards, and try to aggressively convert others who hold different religious beliefs (Brandt & Van Tongeren, 2017). For instance, religious people appear to derogate atheists, and judge them to be immoral, merely because they do not believe in God (Schiavone & Gervais, 2017; Swan & Heesacker, 2012). In one study, Hammer, Cragun, Hwang, and Smith (2012) asked atheists to report their experience of coming out to their religious community about their new belief. Although the authors were not studying blame, the kinds of responses that atheists reported experiencing fit the profile. For instance, 75% of atheists reported being confronted and

told that they were being immoral in their belief, 45% reported being verbally harassed, and 44% reported being asked to give up their atheism. Taken at face value, this work suggests that religious individuals judge atheism to constitute an immoral belief system, and on that basis, blame, criticize, and pressure atheists to recant it.

Belief blame does not appear to be limited to the domains of moral, political, and religious ideology either. As noted above, in our own work, we have observed that people readily blame others for their immoral beliefs, such as the belief held by a scorned child that, "life would be better if my mother does not survive" (see Figure 1). In fact, across all studies in which we measured blame for objectionable moral beliefs (most of which had little to do with political or religious conflicts), participants, on average, attributed substantive blame (Cusimano & Goodwin, 2019).

Recent work also suggests that some people hold others responsible for possessing illogical or irrational factual beliefs. Ståhl, Zaal, and Skitka (2016) asked participants to evaluate a person who forms an irrational belief – such as a belief in homeopathy or astrology. Not all participants negatively evaluated these individuals or their beliefs. However, a large proportion of participants thought that it was morally bad to hold unfounded beliefs like these, and reported that the target was blameworthy (and ought to be punished) for doing so. Thus, preliminary work on how people evaluate others in light of their beliefs suggests that people blame others who hold beliefs – be they moral, political, religious, or factual – that they deem wrong or false.

Blame for ignorance. Finally, many scholars point to cases of unknowingly causing (or failing to prevent) harm (i.e., causing harm as a function of ignorance) as posing a threat to control-based theories. According to control theories, someone is only blameworthy for unknowingly causing harm if the reason they did not know they were causing harm can be traced back to an earlier bad choice (Rosen, 2004; Smith, H., 1983). So, for instance, a parent who hears their young child start the bath but refuses to investigate, is liable for their child drowning even if they are, in that moment, ignorant of what their child is doing. Their ignorance of the threat to their child's life can be traced back to their earlier choice not to investigate – this choice creates a culpable willful ignorance and represents an unjustified decision to disregard the foreseeable risks.

But some cases of ignorance appear blameworthy even when they cannot be traced back to intentional choices in this way (Sher, 2006; Smith, A., 2005; Smith, H., 2011; Vargas, 2005; but see Fischer & Tognazzini, 2009, for a response to these challenges). For instance, people appear blameworthy when they fail to *remember*, such as when someone forgets to call their mother on her birthday or forgets to buy their spouse an anniversary gift (Smith, A., 2005). Relatedly, people appear to be blameworthy when they cause harm by *failing to notice* something that they should have noticed. Smith (2011) and Sher (2006) illustrate this with the imaginary case, *Ryland*, telling an unintentionally hurtful joke:

> Ryland is very self-absorbed. Though not malicious, she is oblivious to the impact that her behavior will have on others. Consequently, she is bewildered when her rambling anecdote about a childless couple, a person with a disability, and a financial failure is not

well received by an audience that includes a childless couple, a person with a disability, and a financial failure.

Smith and Sher suggest that Ryland is blameworthy for upsetting her audience. However, Ryland did not choose to fail to notice that her joke would upset her audience. And, unlike the cases of willful ignorance above, Ryland's ignorance does not stem from some prior conscious choice to ignore information. This inability to trace back Ryland's ignorance to some prior choice raises the question of what does explain her ignorance, and on what basis she is responsible for it.

A likely diagnosis is that Ryland failed to notice that her joke would be upsetting because of the attitudes that she holds, namely, her low degree of care and concern for others. Indeed, one's cares, concerns, and values affect what one attends to and wonders about. Had Ryland been the sort of person who cares more about others' feelings, she would have wondered how her joke would affect this particular audience. This naturally would have led her to attend to who was in the audience, notice that (for instance) a person with a disability was there, spend a moment to think about how the joke would affect that person, and then realize that her joke would offend them. In light of this observation, Sher and Smith have a ready explanation for why Ryland seems blameworthy for causing harm through failing to notice: She is blameworthy for having the wrong kinds of cares, concerns, and values (which explain her failure to notice the harm she was causing). The same logic applies to everyday failures to remember: When someone forgets a birthday or anniversary, it is reasonable to infer that they do not care sufficiently about their friend or spouse – if they did care, they would have remembered.

To our knowledge, little work has directly investigated people's reactions to harms caused by failures to notice. Some preliminary work comes from Murray and colleagues (2019) who gave participants a story about a man who needs to pick up materials for a birthday party on his way home from work. For a variety of reasons, including his getting excited or stressed about some piece of news, he does not notice the store on his drive home, passes it, and so fails to pick up the ingredients. Importantly, participants tended to say that it would be right for his spouse to blame him for forgetting to pick up the ingredients.

Summary

We have reviewed several cases in which people appear to blame others for possessing objectionable mental states. One sort of case involves attitudes, desires, or affective reactions that are outright immoral – such as when someone experiences *schadenfreude* or possesses a desire for harm to occur. Another sort of case involves objectionable beliefs, that are either immoral, politically or religiously incendiary (to a given observer), or simply false. And a still further case involves behaviors that are not unambiguously immoral (e.g., forgetting a birthday), but that seem blameworthy because they reveal objectionable mental states. Though experimental data is sparse, the emerging evidence does vindicate philosophers' speculations – people often do blame others for these kinds of objectionable mental states, suggesting that this sort of blame is part and parcel of our ordinary social practices.

This naturally then raises the question of *why* people blame others for mental states. According to theorists who regard mental state blame as posing a challenge to control-based theories, people blame others in these situations despite thinking that the mental state in question is mostly outside the control of the person who holds it. If this is true, then blame is being assigned simply because a person possesses bad dispositions (or character). However, it is not a given that this is what ordinary people are doing, and until recently, there was little evidence speaking to this question. An alternative perspective, contra to that held by philosophers who hold up these examples as undermining control theories, is that ordinary people regard the objectionable mental states in these examples as quite controllable and assign blame in proportion to their judged controllability. If this is true, then these cases of everyday blame would not constitute a threat to control-based theories, and would in fact bolster such theories. As we review below, this is indeed what we have observed in recent research. People attribute control to others over their mental states (objectionable and otherwise), and assign blame in line with these judgments of control.

Empirically investigating everyday mental state evaluation

Empirical data on attributions of control

When asked directly, people tend to say others have some degree of intentional control over their mental states. For instance, in one study we recently conducted (Cusimano & Goodwin, 2019), we asked college undergraduates to write down the first emotions, beliefs, desires, and attitudes that they could think of. College students nominated the sorts of attitudes one would expect them to encounter and evaluate in their everyday life, including believing in God, thinking that one is in a bad relationship, wanting to own a car, feeling anxious about an exam, or respecting or appreciating a professor. We then gave the most common mental states to a separate group of students and had them rate how much control they thought the holder of the relevant mental state had. We had students make similar judgments for obviously voluntary and controllable behaviors, like *speaking* or *avoiding someone*, and for obviously involuntary and uncontrollable behaviors like *sneezing* or *fainting*. The key finding was that students thought that most mental states were moderately controllable (Figure 2). That is, when judging whether someone "intentionally chose" to hold a mental state or could "choose to stop" holding that mental state if they wanted to, people's ratings of all kinds of mental states fell somewhere in between the very high control attributed to voluntary behaviors and the very low control attributed to involuntary reactions.

The robustness of these results is striking. For instance, one natural worry is that people are not fully considering what it is like to encounter and think about others' mental states, and so their judgments reflect detached, abstract judgments. But, in a follow-up study, the same finding held even when we made these judgments more concrete. Participants nominated a bad belief, emotion, desire, or attitude that they had recently attributed to someone they knew well. They then wrote out in detail everything they could remember about the circumstances in which they learned about that mental state and why they think it occurred. Even after participants had

brought this mental state to mind in a highly specific way, they reported that their friend intentionally chose to adopt the attitude, and that they could have chosen to stop holding the attitude if they had wanted to. Numerous studies across different labs, studying different contexts and using different measures of control, now show that people readily attribute control to others over a wide variety of mental states (Cusimano & Goodwin, 2019, 2020, 2021; Cusimano & Lombrozo, 2021; Turri et al., 2018; Weis et al., 2021).

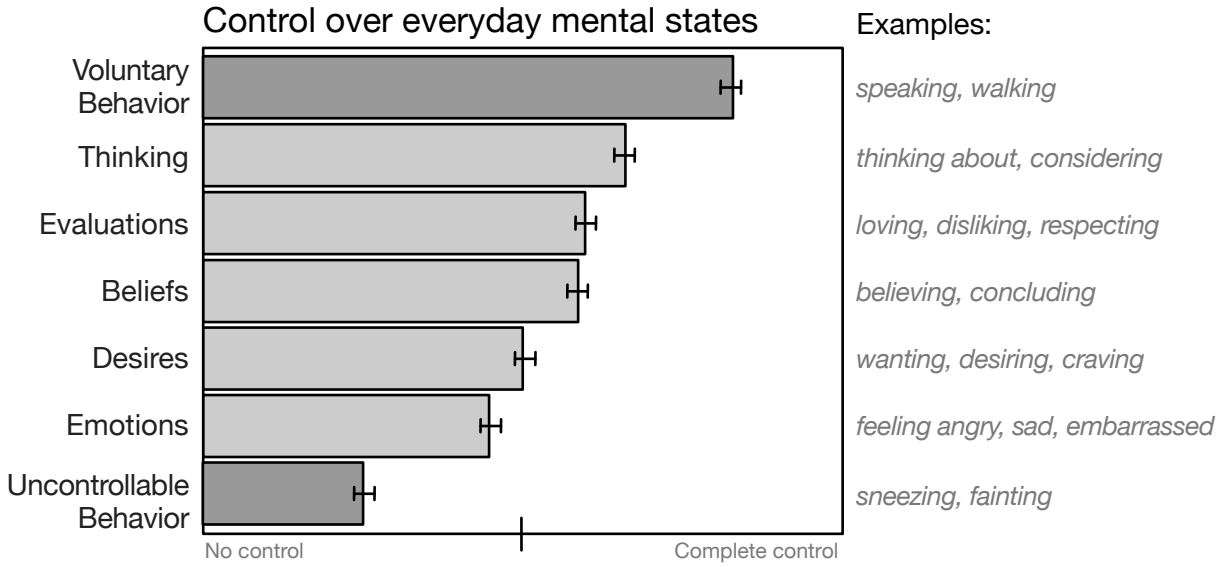

Figure 2. Attributions of control for mundane involuntary behaviors (e.g., sneezing), mental states (light gray bars), and prototypical voluntary behaviors (e.g., speaking).

It is important to note that people do not usually attribute the same degree of control for mental states as they do for simple, voluntary actions such has speaking. In this sense, it seems that lay people largely agree with philosophers think that mental states are not directly changeable through simple acts of will. Rather, a more accurate characterization is that people think others have *effective control* over their mental states, often by indirect means.[2] That is, people intuitively think that, if someone *wanted* to adopt or drop a mental state, then they would be able to find a way to do so. Thus, insofar as lay people seem to hold different views about mental state control than philosophers who claim that mental states are uncontrollable, that discrepancy reflects different judgments about how effectively people can intentionally change their minds.

[2] Emerging work from our lab has identified some of the ways that people think others exercise indirect control over their attitudes. Some strategies we have identified include rationally re-evaluating one's current attitude (Cusimano & Goodwin, 2021) and directing one's attention toward or away from, reasons that favor certain conclusions (Cusimano, Zorrilla, Danks, & Lombrozo, 2021). However, research on this question is still in its early stages.

In theory, people might regard mental states as controllable without also gauging blame for objectionable mental states on the basis of their perceived controllability. If this were true, it would be a mark against control theories, which state that control is a crucial input to blame. However, this in principle worry deflates in practice: Studies that have measured both control and responsibility (for mundane mental states) or both control and blame (for objectionable mental states) have documented a tight correspondence between the two.[3]

Empirical data on mental state blame and responsibility

Recent work reveals that people differ in their intuitions about how controllable others' mental states are. This individual variation provides one source of evidence regarding the role of control in people's assessments of blameworthiness and attributions of responsibility. In our work, people who tended to say that a mental state was uncontrollable also tended to say that the person holding or experiencing that state was not responsible (or blameworthy) for it (Cusimano & Goodwin, 2019). In fact, almost no participants in our studies appear to endorse both that (a) that emotions, beliefs, desires, and other attitudes are uncontrollable, and (b) that others are blameworthy for them (see Figure 3a). We observed the same pattern of results when participants attributed responsibility for everyday unobjectionable mental states (Figure 3b). Individual differences of this sort can also be observed across different cultural or religious groups. Cohen and Rozin (2001) found that, compared to Jews, Christians (and especially Catholics) tended to judge the mind as more controllable. Commensurate with these differences, Catholics also tended to express more moral outrage towards others who held immoral thoughts (like thinking adulterous thoughts or feeling antipathy toward one's parents), further supporting a clear connection between control and blame.

[3] Though there are nuanced differences between attributions of blameworthiness and responsibility, we will present data from experiments that study either of these judgments in relation to mental states.

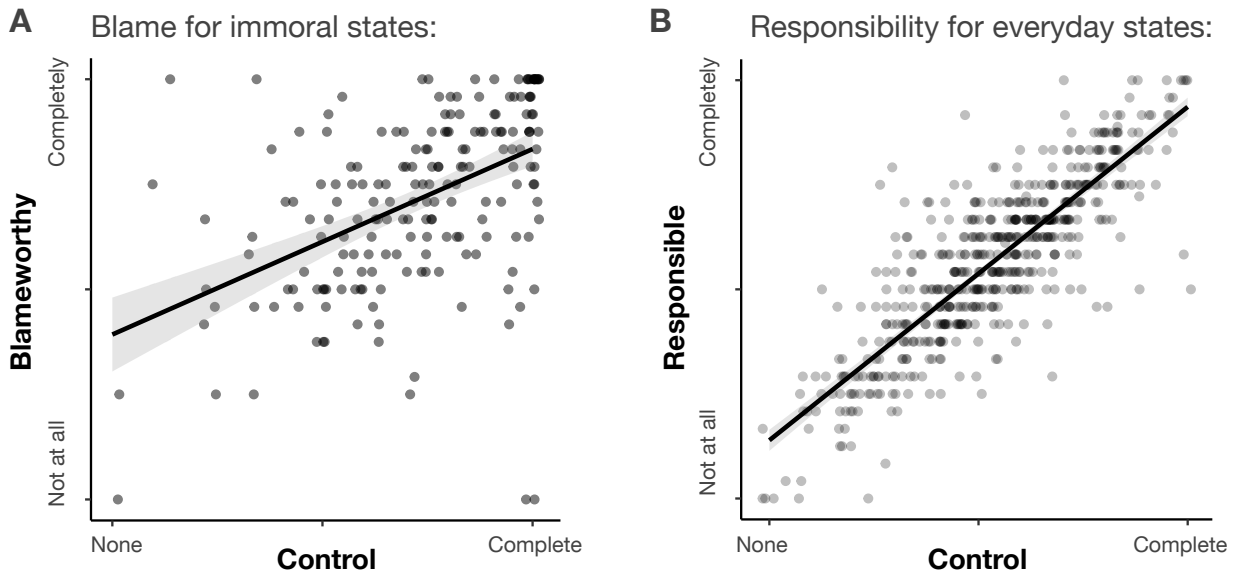

Figure 3. The relationship between judgments of control, blameworthiness, and responsibility. (A) The relationship between control and blameworthiness in Cusimano and Goodwin (2019) Study 4. See Figure 1 for average control ratings. (B) The relationship between control and responsibility ratings in Cusimano and Goodwin (2019) Study 1, based on average control judgments for emotions, desires, beliefs, and evaluations. See Figure 2 for average control ratings for these mental states.

Some of the early empirical work showing support for the role of control in everyday moral judgment did so by showing that people hold others less responsible for physical conditions (such as being overweight or having a disease or disability), when they think that those conditions are the product of uncontrollable genetic or biological forces (Weiner, 1995). Researchers have found the same reasoning applies to mental traits, such as sexual orientation and chronic mental illness. For instance, early research on homophobia found 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). Similarly, dozens of studies show that people who attribute trait-like mental disorders like depression or schizophrenia to uncontrollable causes – such as, again, genetics or biological disfunction – attribute less blame and responsibility to persons who suffer from those disorders (Ahn, Kim, & Lebowitz, 2017; Haslam & Kvaale, 2015).

We recently investigated whether explaining immoral mental states with biological causes similarly reduces perceived control over the attitude (Cusimano, 2019). In one study, participants read about someone who held an immoral mental state, like wanting his mother to die while she is recovering in the hospital from a car crash (discussed above). Participants also learned that the person who held the attitude had a developmental disorder that was either unrelated to the feeling (e.g., he has difficulty reading) or that plausibly explained his feeling

(e.g., he lacks the ability to feel empathy for others). As expected, when there was a plausible, uncontrollable biological cause of the immoral feeling, people judged the person to have less control over whether they experienced it. Most importantly, the biological explanation selectively affected how morally blameworthy people judged the target to be. Specifically, participants blamed him less for his attitude, and reported being less willing to criticize him, in the biological cause condition relative to the control condition. However, they rated him as possessing equally poor character, and expressed an equally strong desire to stay away from him, regardless of whether or not the immoral attitude had a biological explanation.

Concurrent research in emotion regulation demonstrates that attributions of control play an important role in how people relate to their own and others' emotions. Many studies now show that people who believe that emotions are controllable are more likely to try and regulate their emotions (Ford & Gross, 2019), and more likely to react negatively to their own unwanted emotions (Mitmansgruber, Beck, Höfer, & Schüßler, 2009). For instance, in one study, people who thought that they ought to be able to control their emotions were more likely to get angry at themselves for episodes of unwanted emotionality (Mitmansgruber et al., 2009). Attributions of emotion control also predict people's attitudes towards other's emotions. People who tend to think emotions are more controllable also tend to react less supportively towards others who experience negative emotions (Cusimano & Goodwin, 2021; Halberstadt et al., 2013; Tullet & Plaks, 2016). They also are more likely to react with anger or frustration towards others they view as capable, but unwilling, of controlling their emotions. In one study, we had participants report a recent time in their life in which someone close to them had a strong emotional reaction, discuss the circumstances surrounding the emotion, and then report what they thought and how they reacted (Cusimano & Goodwin, 2021). Consistent with the data above, many participants reported that they had criticized their close other for their emotion, expressed irritation at them for feeling this emotion, or that they had refused to help or even feel sorry for their close other. Strikingly, the most robust predictor of people's unsupportive reactions was their judgment that their close other could have chosen to stop feeling bad if they had wanted to.

Summary

A careful examination of how people attribute blame for mental states reveals the familiar signature of control-driven moral evaluation. People attribute at least some, and sometimes relatively high, control to others over the majority of their mental states in everyday life. Specifically, people regard others' mental states as (mostly) intentionally chosen, and believe that others can (mostly) change or drop those mental states if they wish to. The amount of control people attribute predicts how responsible and blameworthy they judge others to be for having those mental states. We reviewed three pieces of evidence in support of this latter claim. First, individual differences in the perceived controllability of specific mental states predict judgments of responsibility and, in the case of objectionable mental states, blameworthiness for them. This pattern replicates across mental state type (e.g., emotions, beliefs, desires, etc.), across studies and experimental contexts, and across both morally charged and morally neutral

mental states. Second, when people think that objectionable mental states are caused by uncontrollable forces, such as a person's genes, they judge others with those states as less blameworthy for them. Third, and finally, we noted a special case of mental state responsibility, namely, responsibility for emotions. Emerging work in this domain shows that people who view emotions as controllable seem to hold themselves and others more responsible for regulating them. And mirroring the other work above, people who think that others can control their emotions tend to react negatively when they fail to do so. Thus, while the evidence is preliminary – a point we address below – it appears that the question of why people attribute moral responsibility for mental states favors a control-based answer.

What do these data entail for the philosophical question of mental state responsibility? Insofar as theories of moral responsibility attempt to capture the implicit commitments of everyday moral judgment, philosophers should strongly favor control-based theories. As noted above, mental state responsibility constituted an ideal scenario in which people might plausibly blame others for things that they consider outside of those others' control. Yet, while people do indeed blame others for holding inherently immoral mental states, they do not appear to do so without reference to control. Instead, they attribute control to others over their objectionable mental states and blame them on this basis. Although the empirical picture remains incomplete, the burden is now on philosophers who want to challenge control theories to document how the apparent evidence for control-based theories is misleading. In the remainder of this chapter we critically examine two strategies that philosophers and psychologists might take towards this end.

Objections and Future Directions

Objection 1: The evidence is confounded

The best argument against the findings above is that attributions of control are a proxy for some other judgment that people actually care about when attributing moral responsibility. This objection shares the same structure as the classic "third variable" problem in experimental psychology: that the measured variable (in this case, control) covaries with some other variable (unspecified), which is ultimately the cause of people holding others responsible for mental states. Indeed, the most prominent weakness of the empirical literature surveyed above is that it has not yet dissociated perceived mental state control from other variables that have been proposed as key drivers of moral responsibility (and that do not necessitate control). Candidate alternative variables include the degree to which a given mental state is open to rational re-evaluation (Smith, A., 2008), is reflective of a person's "deep self" (Sripada, 2017), or is reflective of a stable and important component of a person's moral character (Bayles, 1982; Hieronymi, 2008; Smith, H., 2011). As philosophers have noted, many of the most plausible candidates for moral responsibility – attributions of control, the deep self, rational re-evaluation, etc. – tend to make the same predictions about responsibility, which complicates the task of discerning the true explanation. Mental state blame has often been seen as a promising way to distinguish these theories; however, the strategy of using commonplace mental state blame to

support character-based theories, and to rule out control-based theories no longer seems tenable. Future work will have to find new ways to dissociate these theories.

The idea that control is merely a proxy for some other judgment also cuts in the other direction, revealing a potential source of concern for proponents of alternative theories. That is, just as it is possible that control judgments act as a mere proxy for these alternatives, it is also possible that alternatives accounts covertly import intuitions about mental state control. Consider for example Smith, A. (2005)'s proposal that people are responsible for mental states that are in principle open to modification through rational re-evaluation (e.g., a racist belief) but not states that are less integrated with someone's rational capacities (e.g., a simple phobia). Based on the work reported above, people are also likely to regard the former mental state as more controllable than the latter, precisely because they treat rational re-evaluation as one way of intentionally (albeit indirectly) changing one's mind. As suggested by Smith, H. (2011), Smith A.'s proposal may seem to make reasonable predictions about mental state responsibility, but only because it leverages intuitions about control. A strong defense of the claim that people hold others responsible for mental states (or other behaviors) despite how (un)controllable they are requires properly dissociating the relevant predictors, such that it clear that the alternative criterion drives responsibility and blame in the absence of control.

Objection 2: The evidence is incomplete

Philosophers have been discussing the question of mental state responsibility for much longer than psychologists have been directly investigating it. So, it is no surprise that the set of theoretical challenges to control outstrips the set of cases that have been studied empirically. It may very well be that, in certain situations that are yet to be examined, people do blame others for conduct that they regard as uncontrollable, which would show that control-based theories fail to account completely for our everyday moral judgment. Two cases in particular might fulfill this possibility: blame for hurtful attitudes and blame for ignorance. It is possible that people blame others for failing to be loving partners, proud fathers, respectful children, and so on, without regard to whether those persons could have chosen to have better attitudes. Likewise, it is possible that people blame others for failures to notice or remember without reference to control, simply because those failures indicate uncaring attitudes. The resolution of this issue awaits further research. However, based on the apparent importance of control in other domains of mental state evaluation, we believe it is unlikely that research in these remaining domains will point towards a radically different pattern of judgments.

Concluding Remarks

Decades of work in empirical psychology has supported control-based theories of moral responsibility. The ubiquity of control in underpinning everyday moral judgments has in turn been a source of motivation for normative theories of moral responsibility and blame. However, the role of control in descriptive and normative theories of moral responsibility has not gone unchallenged. One prominent challenge is that people seem to blame others simply for holding

objectionable mental states, and that such mental states are judged to be uncontrollable. However, recent empirical work on the everyday practice of mental state blame casts this challenge in a new light. Partly vindicating this challenge, evidence suggests that people do routinely blame others for their bad mental states. However, people also attribute considerable control to others over their mental states, with judgments of control predicting subsequent judgments of blame and responsibility. This second set of findings undermines the threat posed to control-based theories by the mental state challenge, and instead appears to broaden the application of such theories. And because mental state blame represented the best case against control-based theories, such theories are likely here to stay.

References

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

Ahn, W., Kim, N. S., & Lebowitz, M. S. (2017). The role of causal knowledge in reasoning about mental disorders. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (p. 603–617). Oxford University Press.

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

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

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

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

Brandt, M. J., & Van Tongeren, D. R. (2017). People both high and low on religious fundamentalism are prejudiced toward dissimilar groups. *Journal of Personality and Social Psychology, 112*, 76-97.

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science, 23*(1), 27-34.

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

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

Crawford, J. T., Brandt, M. J., Inbar, Y., Chambers, J. R., & Motyl, M. (2017). Social and economic ideologies differentially predict prejudice across the political spectrum, but social issues are most divisive. *Journal of Personality and Social Psychology, 112,* 383-412.

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

Cusimano, C. (2019) *Attributions of mental state control: Causes and consequences* (Publication No. AAI22588322). [Doctoral dissertation, University of Pennsylvania]. Dissertations available from ProQuest.

Cusimano, C., & Goodwin, G.P. (2019). Lay beliefs about the controllability of everyday mental states. *Journal of Experimental Psychology: General, 148*(10), 1701-1732.

Cusimano, C., & Goodwin, G.P. (2020). People judge others to have more voluntary control over beliefs than they themselves do. *Journal of Personality and Social Psychology, 119*, 999-1029.

Cusimano, C., & Goodwin, G.P. (2021) People regulate each other's emotion regulation. PsyArXiv. https://doi.org/10.31234/osf.io/abq3v

Cusimano, C., & Lombrozo, T. (2021). Morality justifies motivated reasoning in the folk ethics of belief. *Cognition.* 104513.

Cusimano, C., Zorrilla, N., Danks, D., & Lombrozo, T. (2021). Reason-based constraint in theory of mind. *Proceedings of the 43nd Annual Conference of the Cognitive Science Society,* Cognitive Science Society.

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives, 30*(3), 133-40.

Fincham, F. D., & Jaspers, J. (1979). Attribution of responsibility to the self and other in children and adults. *Journal of Personality and Social Psychology, 37*(9), 1589–1602.

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

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

Fischer, J. M., & Tognazzini, N. A. (2009). The truth about tracing. *Noûs, 43*(3), 531-556.

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

Garrett, K. N., & Bankert, A. (2018). The moral roots of partisan division: How moral conviction heightens affective polarization. *British Journal of Political Science*, 1–20.

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*(8), 1077-1091.

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

Haider-Markel, D. P., & Joslyn, M. R. (2008). Beliefs about the origins of homosexuality and support for gay rights: An empirical test of attribution theory. *Public opinion quarterly, 72*(2), 291-310.

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

Halberstadt, A. G., Dunsmore, J. C., Bryant, A., Parker, A. E., Beale, K. S., & Thompson, J. A. (2013). Development and validation of the parents' beliefs about children's emotions questionnaire. Psychological Assessment, 25(4), 1195-1210.

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

Haslam, N., & Kvaale, E. P. (2015). Biogenetic explanations of mental disorder: The mixed-blessings model. *Current Directions in Psychological Science, 24*(5), 399-404.

Hieronymi, P. (2008). Responsibility for believing. *Synthese, 161*(3), 357-373.

Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition, 182*, 331-348.

Knobe, J., & Doris, J. M. (2010). Responsibility. In J. M. Doris (Ed.) *The moral psychology handbook* (p. 321–354). Oxford University Press.

Levy, N. (2005). The good, the bad, and the blameworthy. *Journal of Ethics and Social Philosophy*, *1*(2), 1-16.

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

Markman, K. D., & Tetlock, P. E. (2000). 'I couldn't have known': accountability, foreseeability and counterfactual denials of responsibility. *British Journal of Social Psychology*, *39*(3), 313-325.

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

McGrory, M. (1988, November 10). Deadly seriousness. *The Washington Post*, pp. A38.

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*(4), 448-453.

Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology, 116*, 215-236.

Murray, S., Murray, E. D., Stewart, G., Sinnott-Armstrong, W., & De Brigard, F. (2019). Responsibility for forgetting. *Philosophical Studies*, *176*(5), 1177-1201.

Nelkin, D. (2011). *Making sense of freedom and responsibility*. Oxford University Press.

Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil*., Herzliya series on personality and social psychology (pp. 91–108). Washington, DC US: American Psychological Association.

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

Robbennolt, J. K. (2000). Outcome severity and judgments of "responsibility": A meta-analytic review. *Journal of Applied Social Psychology, 30*, 2575–2609.

Rosen, G. (2004). Skepticism about moral responsibility. *Philosophical perspectives, 18*, *Ethics*, 295-313.

Royzman, E., & Kumar, R. (2004). Is consequential luck morally inconsequential? Empirical psychology and the reassessment of moral luck. *Ratio*, *17*(3), 329-344.

Ryan, T. J. (2014). Reconsidering moral issues in politics. *The Journal of Politics, 76*(2), 380-397.

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

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

Schiavone, S. R., & Gervais, W. M. (2017). Atheists. *Social and Personality Psychology Compass*, *11*(12).

Sher, G. (2006) Out of control. *Ethics, 116*, 285–301

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

Shultz, T. R., Wright, K., & Schleifer, M. (1986). Assignment of moral responsibility and punishment. *Child Development, 57*(1), 177–184.

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

Smith, A. M. (2005) Responsibility for attitudes: Activity and passivity in mental life. *Ethics* 115, 236–271

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

Smith, H. (1983). Culpable ignorance. *The Philosophical Review, 92*(4), 543-571.

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

Sripada, C. (2017) Frankfurt's unwilling and willing addicts. *Mind, 126*, 781–815

Ståhl, T., Zaal, M. P., & Skitka, L. J. (2016). Moralized rationality: Relying on logic and evidence in the formation and evaluation of belief can be seen as a moral issue. *PloS one*, *11*(11), e0166332.

Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy, 48,* 187–211.

Swan, L. K., & Heesacker, M. (2012). Anti-atheist bias in the United States: Testing two critical assumptions. *Secularism and Nonreligion, 1*, 32-42.

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

Tullett, A. M., & Plaks, J. E. (2016). Testing the link between empathy and lay theories of happiness. *Personality and Social Psychology Bulletin*, *42*(11), 1505-1521.

Turri, J., Rose, D., & Buckwalter, W. (2018). Choosing and refusing: Doxastic voluntarism and folk psychology. *Philosophical Studies, 175*, 2507–2537.

Vargas, M. (2005). The trouble with tracing. *Midwest studies in philosophy, 29*, 269-291.

Wallace, R. J. (1994). *Responsibility and the moral sentiments*. Harvard University Press.

Walster, E. (1966). Assignment of responsibility for an accident. *Journal of Personality and Social Psychology, 3*, 73–79.

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

Weiner, B., Amirkhan, J., Folkes, V. S., & Verette, J. A. (1987). An attributional analysis of excuse giving: Studies of a naive theory of emotion. *Journal of personality and social psychology*, *52*(2), 316-324.

Weiner, B., Figueroa-Munioz, A., & Kakihara, C. (1991). The goals of excuses and communication strategies related to causal perceptions. *Personality and Social Psychology Bulletin*, *17*(1), 4-13.

Weiss, A., Forstmann, M., Burgmer, P., & Weiss, A. (2021). Moralizing mental states: The role of trait self-control and control perceptions. *Cognition*.

Wolf, S. (1990). *Freedom within reason.* Oxford University Press.