

Do moral beliefs motivate action?

Rodrigo Díaz, Centre for Research in Ethics, University of Montreal

(forthcoming in Ethical Theory and Moral Practice, please cite the published version)

Abstract

Do moral beliefs motivate action? To answer this question, extant arguments have considered hypothetical cases of association (dissociation) between agents' moral beliefs and actions. In this paper, I argue that this approach can be improved by studying people's actual moral beliefs and actions using empirical research methods. I present three new studies showing that, when the stakes are high, associations between participants' moral beliefs and actions are actually explained by co-occurring but independent moral emotions. These findings suggest that moral beliefs themselves have little or no motivational force, supporting the Humean picture of moral motivation.

Keywords

Moral motivation; Moral beliefs; Moral emotions; Humean theory of motivation; Sentimentalism; Experimental philosophy

Acknowledgments

This work is funded by the Swiss National Science Foundation (research project P2BEP1_200040). I would like to thank Jules Salomone-Sehr and Christine Tappolet for their comments on early drafts of the paper.

Ethics information

This research was approved by the University of Montreal Research Ethics Committee (Protocol ID: CERSC-2021-008-D). Participants provided informed consent and were compensated for their participation in the studies.

Data availability

Data and materials are available at https://osf.io/hxp46/?view_only=43dfdfdc1d03449f8fd52f7ffe820875

1. Introduction

Imagine you are a parent, and your daughter has torn out the pages from a rare edition of your favorite novel. You are quite angry and the thought of physically punishing your daughter crosses your mind. Nevertheless, you refrain from doing so. What has motivated your self-control? Is it your belief that parental violence is morally wrong, or your aversion towards it?

According to Humean views, moral motivation always originates in desire (Smith 1987; Sinhababu 2009; Blackburn 1998; Prinz 2007). On such views, beliefs explain our conduct only to the extent that they help us satisfy our desires, or determine the means required to achieve our ends. As Hume himself famously put it, “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (Hume, 1739, §2.3.3). If this is right, your belief that parental

violence is wrong is motivationally inert without pre-existing desires or aversions, the same way as your belief that there is food in the fridge will not prompt you to eat it unless you are hungry or whimsical.

Against Motivational Humeanism, some have posited that moral beliefs (perhaps unlike other kinds of beliefs)¹ are enough to motivate action. In other words, we can be motivated to do something *just because* we believe it is the right thing to do (May 2013; Shafer-Landau 2003). On these Anti-Humean views, moral beliefs motivate without the help of antecedent desires, either because moral beliefs are intrinsically motivating (Dancy 1993) or because they generate motivational states (Nagel 1970; Darwall 1983). Descriptive beliefs such as the belief that there is food in the fridge might be motivationally inert. Anti-Humean will insist, however, that your belief that parental violence is morally wrong suffices to motivate action.

There is much at stake in the debate between Humean and Anti-Humean views of moral motivation. Most notably, claims about the motivational force of moral beliefs play a key role in the so-called Moral Problem (Smith 1994). The Moral Problem arises due to the combination of three plausible claims: (1) Beliefs themselves do not motivate or generate motivational states (Motivational Humeanism), (2) Sincerely holding a moral judgment motivates to act in accordance with it (Judgment Internalism), and (3) Moral judgments are beliefs (Moral Cognitivism). If all three claims are true, there are no moral judgments. To avoid this conclusion, we have to reject one of the three claims in the puzzle, and Motivational Humeanism is sometimes thought to be the weakest link.

In this paper, I provide an empirical assessment of Motivational Humeanism in the moral domain. Both advocates and detractors of Motivational Humeanism have taken

¹ Note that not every Anti-Humean view gives a special status to moral vs. non-moral beliefs, and some even advise against this distinction (Bromwich 2010). The discussion in this paper is restricted to the moral domain, and will thus not tackle this issue.

it to involve an empirical claim about the motivational force of moral beliefs (May 2018; Sinhababu 2009; Sobel 2013; but see Arruda 2017). However, and to my knowledge, no study to date has empirically tested this claim.² In the following, I will motivate this investigation by arguing that empirical research methods can help us address two important limitations of extant arguments building on hypothetical cases. Afterwards, I will present the results of three original studies testing the relation between participants moral beliefs, emotions and actions, and discuss their theoretical implications.

2. Hypothetical cases of moral motivation

Are moral belief sufficient to motivate action? Changes in moral beliefs seem to reliably cause behavioral changes, suggesting that moral beliefs indeed push us to act. For example, someone who comes to believe that eating meat is morally wrong (after thinking otherwise) will probably reduce their meat consumption. However, it seems like agents with the same moral beliefs can be differently motivated. Indeed, not all those who believe that meat-eating is morally problematic might stop eating meat, suggesting that moral beliefs are motivationally inert. These two examples present the structure of extant case-based arguments for and against Motivational Humeanism.³ There are at least two types of cases that are relevant in this context, which we will discuss in turn.

2.1. Cases of association between moral beliefs and action

² Previous work has used empirical evidence to test Judgment Internalism (Roskies 2003), and Moral Cognitivism (Prinz, 2007). This work has important connections with the investigations presented in this paper, both in terms of method and content.

³ Motivational Humeanism has been assessed using different types of considerations. For example, some authors have considered questions regarding mental states' "direction of fit" (Smith 1987; 1994; Little 1997; Price 1989), while others have looked at the phenomenology of moral motivation (Schurman 1894; May 2013; Shafer-Landau 2003). Here, I will focus on arguments building on hypothetical cases.

The first type of case shows an association between moral beliefs and action. More specifically, it shows that changes in moral beliefs are associated with changes in one's actions. This suggests that moral beliefs are sufficient to motivate action, supporting Motivational Anti-Humeanism. A much-discussed case of this type is the following:

Roberta grows up comfortably in a small town. [...] She is aware that there is poverty and suffering somewhere, but sees no relation between it and her own life. On going to a university she sees a film that vividly presents the plight of textile workers in the southern United States: the high incidence of brown lung, low wages, and long history of employers undermining attempts of workers to organize a union, both violently and through other extralegal means. Roberta is shocked and dismayed by the suffering she sees. After the film there is a discussion of what the students might do to help alleviate the situation. It is suggested that they might actively work in promoting a boycott of the goods of one company [...] She decides to donate a few hours a week to distributing leaflets at local stores. (Darwall 1983, p. 39)

According to Stephen Darwall, who first presents this case, Roberta forms the belief that the situation of these textile workers is unacceptable upon watching the documentary, and this belief motivates her to join the boycott that the students organize. We might also attribute Roberta the desire to help the workers, but this desire would always be subordinated to her becoming aware of the injustice that the workers suffer. In other words, this desire would be belief-dependent or "motivated" (see Nagel, 1970). Thus, Darwall concludes that Roberta's newly formed moral belief about the situation of the workers sufficiently motivates her actions.

Advocates of Motivational Humeanism, however, offer an alternative interpretation of Roberta's case. According to Neil Sihlababu (2009, pp. 483-489), Roberta's behavior

is better explained in terms of a desire to alleviate suffering, a desire that she has before watching the documentary, and explains why she “is shocked and dismayed by the suffering she sees” (Darwall 1983, p. 39). On this reading, Roberta desires to alleviate suffering, and after discovering the suffering of the textile workers in the documentary, she decides to help them. Roberta might also form the belief that the situation of the workers is unjustifiable, but what ultimately motivates her to act is her desire to alleviate suffering.

We of course don't know what really motivates Roberta, as this is a purely hypothetical case. Both Humean and Anti-Humean explanations of Roberta's actions are plausible, and favoring one over the other seems to be largely a matter of our previous theoretical commitments regarding the (lack of) motivational force of moral beliefs.⁴ This way, the discussion reaches an impasse. Arguably, the reason for this impasse is the following:

(Co-occurrence) Associations between moral beliefs and actions (suggest that moral beliefs sufficiently motivate but) can plausibly be explained by co-occurring belief-independent desires.

One could argue that, in order to address Co-occurrence, we can design a hypothetical case in which we stipulate the absence of belief-independent desires. However, if we characterize our cases in a way that rules out this alternative explanation, we would be begging the question against Motivational Humeanism. In particular, if we stipulate that Roberta has no desire to alleviate suffering previous to watching the documentary, this would just be a statement of the Anti-Humean view of moral motivation, and Humeans would no longer consider the case plausible. Thus, it seems

⁴ Proponents of Anti-Humean views argue that there is no need to posit an antecedent desire to explain Roberta's behavior, given that we already have an explanation in terms of her moral beliefs (Darwall, 1983, p. 40). But Humeans argue that there is no need to posit that beliefs have motivational power, given that an explanation in terms of desire is readily available (Sinhababu 2009, p. 487).

like Co-occurrence cannot be addressed by stipulation.

2.2. Cases of dissociation between moral beliefs and action

The second type of case that is used to test moral beliefs' motivational power shows a dissociation between moral beliefs and action. In particular, it shows that agents with the same moral beliefs can nevertheless act differently. This suggests that moral beliefs do not generate motivation, thus supporting Motivational Humeanism. Consider the following case, adapted from Connie Rosati (Rosati 2016, p. 13; see also Shafer-Landau 2003, p. 129):

Luis and Carlos believe that they ought to contribute to famine relief. They are both aware that people all around the world suffer and die from starvation, and they believe that this situation is morally unacceptable. However, Luis donates a percentage of his income to Oxfam, while Carlos does nothing to fight global hunger.

This case nicely fits the Humean framework. Given that both Luis and Carlos act differently but have the same moral beliefs, it must be the case that their motivation is sourced elsewhere. This paves the way for an explanation in terms of desires or emotions. However, as Russ Shafer-Landau has noted, "anti-Humeans needn't deny that desires are motivating states, and so needn't deny that different desires can, even in the face of identical beliefs, alter overall motivation. This admission is compatible with the claim that beliefs can motivate all by themselves." (Shafer-Landau 2003, p. 131).

Here, it is important to note that agents can fail to act because they lack motivation, or because they have countervailing motivations. Given that Anti-Humeans do not deny that states other than belief can motivate, they can explain Carlos' failure to act in terms of countervailing motivations. Indeed, it might be that Carlos does nothing to

fight global hunger because, although his moral beliefs motivate him to take action, he is unwilling to spend money or mistrusts charitable organizations. This interpretation accommodates the case within an Anti-Humean framework, and introduces the second limitation in the use of hypothetical cases:

(Counterbalance) Dissociations between moral beliefs and actions (suggest that moral beliefs themselves do not motivate but) can plausibly be explained by the presence of countervailing motivations.

How can we address this limitation? Once again, one might be tempted to do this by stipulation. We could propose a case in which someone fails to act in accordance with their moral beliefs in the absence of any countervailing motivation. However, Anti-Humeans would not find these cases plausible. As we have seen, Anti-Humeans need to appeal to the presence of countervailing motivations to explain cases in which agents fail to act despite having the relevant moral beliefs. Thus, as it happened with Co-occurrence, it seems like we cannot address Counterbalance by reworking our hypothetical cases.

Note that the limitations listed here are not meant to constitute a comprehensive or fatal criticism of case-based arguments. Instead, the goal of listing these limitations is to identify where the use of empirical research methods can aid extant argumentative practices in philosophy.

3. Empirical data on moral motivation

3.1. The advantages of using empirical research methods

In the previous section, we identified two limitations in the use of hypothetical cases to inform our theories of moral motivation: Co-occurrence and Counterbalance. Here, I argue that both limitations can be addressed by studying actual moral behavior using

empirical research methods.

According to Counterbalance, dissociations between moral beliefs and actions can plausibly be explained by the presence of countervailing motivations. This limitation could be addressed by considering the moral beliefs and actions of not one or two, but hundreds of agents. Consider, for example, the relation between agents' moral beliefs about famine relief and their donations to famine-relief charities. If we consider just two agents, a dissociation between their moral beliefs and actions can be explained by unwillingness to spend money or mistrust of charities. But, if we consider hundreds of agents, some will be unwilling to spend money, but some will be giving. Some will mistrust charities, but others will trust them. Unless these countervailing motivations systematically co-occur with the relevant moral beliefs (e.g., everyone who believe they should donate also tend to mistrust charities), we can be confident that countervailing motivations do not explain a dissociation between moral beliefs and actions, addressing Counterbalance. While considering the beliefs and actions of hundreds of agents would be too demanding for imagination, it is a standard practice in empirical research (Sassenberg and Ditrich 2019).

According to Co-occurrence, associations between moral beliefs and actions can plausibly be explained by co-occurring belief-independent desires. For example, those who believe they ought to fight injustices might also want to alleviate suffering, and those who do not think they ought to fight injustice might also lack the desire to alleviate suffering. If certain beliefs and belief-independent desires tend to co-occur, finding an association between moral beliefs and actions wouldn't show that moral beliefs motivate action, even if we consider hundreds of agents. In order to address this limitation, we have to elucidate the relative contributions of the relevant moral beliefs and desires. As we saw in the previous section, this cannot be done using hypothetical cases. We cannot know, for example, whether Roberta was ultimately

motivated by her beliefs or her desires. However, the use of empirical research methods allows us to measure all potentially explanatory factors (e.g. moral beliefs or desires) and test which one better explains a given outcome (e.g. moral behavior) using statistical tools (see §3.3.).

3.2. Previous studies on moral motivation

A series of extant studies in psychology and economics have shown that individual differences in moral beliefs predict a series of moral-behavioral outcomes. These studies measure a large number of participants' moral beliefs and actions, helping to address Counterbalance. Unfortunately, they do not consider the relative impact of participants' desires. Thus, they do not address Co-Occurrence. In the following, I briefly describe these studies.

Most studies on the relation between moral beliefs and actions have measured participants' moral beliefs using the Moral Foundations Questionnaire (MFQ; Graham et al., 2011), which has been shown to be a reliable and valid measure of moral beliefs (Zhang, Hook, and Johnson 2016).⁵ The idea behind the MFQ is that morality can be built upon a combination of five different foundations: Harm/Care, Fairness/Reciprocity, Ingroup/Loyalty, Authority/Respect, and Purity/Sanctity. The questionnaire consists of a series of items measuring people's moral beliefs regarding these five issues. For example, to measure moral beliefs about Fairness, the MFQ asks to rate statements such as "When the government makes laws, the number one principle should be ensuring that everyone is treated fairly" from 1 ("strongly disagree") to 6 ("strongly agree"), and to measure moral beliefs regarding loyalty, it

⁵ The Moral Foundations Questionnaire is not free of criticism. Most criticisms have focused on whether the questionnaire's items should be divided in the proposed five categories. But the two categories that will concern us in this paper, Harm and Fairness, seem to be adequate (Zakharin and Bates 2021). For discussion regarding the internal consistency of MFQ categories, see Tamul et al. (2020, p. 5-6), but also Graham et al. (2011, p. 370). For an alternative way of dividing the moral domain, see Curry et al. (2019).

asks for agreement with statements such as “People should be loyal to their family members, even when they have done something wrong” using the same scale. Respondents’ answers to these questions provide a numeric measure of (the strength of) their moral beliefs regarding each of the five moral foundations.

Díaz and Cova (2021) found that individual differences in moral beliefs regarding Harm are associated with efforts to adopt health recommendations during the COVID-19 pandemic. In this study, participants reported how many efforts they were making to avoid contact, wash hands regularly, cover their coughs, and stay at home as much as possible, as well as their moral beliefs in the MFQ. Results showed a positive association between moral beliefs regarding Harm (as measured by the MFQ) and efforts to comply with health recommendations. In other words, participants with stronger beliefs about the moral importance of avoiding harm reported more efforts in complying with health recommendations.

Similarly, Schier, Ockenfels, and Hofmann (2016) found that individual differences in moral beliefs regarding Fairness are associated with fair behavior in a classic economic game known as the Dictator Game. The Dictator Game involves two players. One player is allocated a number of resources (e.g. 10 dollars) and is given the opportunity to share some of them with the other player. The second player has no option but to accept the offer, hence the name of the game. The number of resources shared by the “dictator” is taken as a measure of fairness. Schier, Ockenfels, and Hofmann’s study found that participants’ moral beliefs regarding Fairness (as measured by the MFQ) were associated with the percentage of resources they shared in a Dictator Game, such that stronger moral beliefs about fairness were related to fairer allocations.

Note that these studies assume that there is a connection between participants’ general moral beliefs about harm avoidance or fairness, and their moral beliefs about specific acts of avoiding harm (by following health recommendations) or being fair (in

distributing raffle tickets). The fact that they found a significant effect of general moral beliefs on these specific behaviors suggests that this assumption is well-placed. In other words, it seems like participants considered that following health recommendations and sharing raffle tickets are instances of avoiding harm and being fair, respectively.

One might think of these studies as analogous to Roberta's case. As the studies above, Roberta's case shows that different moral beliefs (pre- and post- watching the documentary) are associated with different actions (living a congenial life vs. joining a boycott). The reviewed studies found that, across individuals, different moral beliefs are associated with different actions. However, as it happened in Roberta's case, this association can plausibly be explained in terms of co-occurring but independent desires (or Co-Occurrence). Because extant studies didn't record participants' desires, we cannot rule out the possibility that co-occurring but independent desires explain their results. Studies testing the motivational efficacy of philosophical arguments (Buckland et al. 2021; Lindauer et al. 2020; Schwitzgebel, Cokelet, and Singer 2020; 2021), which arguably aim at changing people's beliefs, face a similar limitation as they also do not measure participants' desires.

3.3. The present studies

A straightforward way to address Co-occurrence would be to rerun the empirical studies described in the previous section including measures of both moral beliefs *and* desires. However, we might better consider the relative contributions of moral beliefs and moral emotions. It is usual among Humeans to contend that moral judgments express or contain emotions, where emotions are understood as closely tied⁶ to desires

⁶ There are different ways of understanding the connection between emotions and desires. For example, emotions can be understood as modulating (Yip 2021), generating (Tappolet 2016), or depending on desires (Schroeder 2004).

(Sinhababu 2017; Blackburn 1998; Prinz 2015).⁷ In particular, much work has been devoted to study “moral emotions” such as outrage, disgust, guilt, admiration, and gratitude (Haidt 2003; Prinz 2011; Tappolet 2016; Cova, Deonna, and Sander 2015). These emotions tend to arise in response to morally significant events such as stealing or giving to charity (Prinz and Nichols 2010), and significantly impact moral behavior (Tangney, Stuewig, and Mashek 2007). Thus, they are a good candidate to explain people’s moral choices, either directly or due to their tight connection to desires.

Before going further, it is important to note that the debate between Humeans and Anti-Humeans builds on the assumption that beliefs and desires are distinct psychological states. However, some questioned this way of carving up the mind, either by positing “besire” states that have both belief-like and desire-like features (Little 1997; Swartzler 2013), or by defending that desires are beliefs about what’s good or ought to be done (Price 1989; Gregory 2017). Here, I assume that there is a distinction between beliefs and desires, and that expressions such as “I believe that stealing is wrong” and “I feel outraged when I steal” refer to two different things.

The present studies will measure participants’ moral beliefs, moral emotions, and moral choices. Measuring both moral beliefs and moral emotions will allow us to test which one better explains moral behavior by using multiple regression statistical analyses (Field 2013; Allison 1999). Multiple regression differs from correlation in that it calculates the relative association between changes in each of two or more independent variables (in our case, moral beliefs and moral emotions) and changes in the dependent variable (in our case, moral behavior) while holding the other

⁷ Consider again the Moral Problem. By rejecting the Humean theory of Motivation, Anti-Humeans can keep the other two claims in the puzzle: Moral Cognitivism, according to which moral judgments are beliefs, and Judgment Internalism, according to which moral judgments themselves motivate action. But Humeans are pressured to reject one of those. In order to solve the Moral Problem, most Humeans reject Moral Cognitivism and keep Judgment Internalism (but see, e.g. Kumar 2016). If we combine the claim that moral judgments express or contain emotions (anti-Cognitivism), and the claim that moral judgments motivate action (Internalism), we have the claim that emotions motivate action.

independent variables constant. This allows us to test which of the independent variables better explains changes in the dependent variable.

Multiple regression can help us elucidate whether a relationship between two variables is explained by a third variable. For example, a criminologist who finds an association between educational programs and low recidivism “may perform a multiple regression to see if this apparent relationship is real or if it could be explained away by the fact that the prisoners who enroll in educational programs tend to be those with less serious criminal histories.” (Allison 1999, p. 2). In our case, multiple regressions will test whether observed associations between moral beliefs and moral choices can be explained away by co-occurring but independent moral emotions. This will address Co-Occurrence (see §3.1).

One might worry that, in order to provide an alternative explanation for the association between moral beliefs and actions, moral emotions have to be independent of moral beliefs. If moral emotions completely depend on moral beliefs,⁸ we cannot test their relative contributions to moral motivation. In statistical terms, this is called multicollinearity, and it happens when two independent variables are almost perfectly associated, i.e. when an increment of ± 1 in one of the variables is associated with an increment of ± 1 in the other variable. Multicollinearity indeed threatens the validity of multiple regression results. Thankfully, there are ways of testing for it. The results of these tests will not only show whether our multiple regression tests are valid, but also whether moral emotions can be distinguished from moral beliefs even if they tend to co-occur.

Data and materials for all studies can be found at

⁸ There are some principled reasons for this worry. For example, some have claim that beliefs about value are necessary and sufficient for emotion (Nussbaum 2001). Note that, as it happened with some theoretical proposals about the nature of desire, this might blur the distinction that grounds the debate between Humean and Anti-Humean views of moral motivation.

https://osf.io/hxp46/?view_only=43dfdfdc1d03449f8fd52f7ffe820875.

4. Study 1: Health behaviors

Study 1 tests whether the previously-recorded association between moral beliefs regarding Harm (as measured by the MFQ) and compliance with official health recommendations during the COVID-19 pandemic (Díaz and Cova 2022) can be explained in terms of co-occurring and belief-independent moral emotions towards Harm.

4.1. Methods

197 participants were recruited through Amazon Mechanical Turk (MTurk) and completed the survey for a monetary payment. 4 did not complete the survey, and 23 failed the attention check (see below), leaving a final sample of 170 participants (66 female, Mean Age = 37.87, SD = 11.967, age-range 20 – 72).⁹

Participants were presented with a series of four recommendations to avoid the spread of the coronavirus: (1) avoid close contact, (2) wash hands often, (3) cover coughs and sneezes, and (4) use a face cover when around others. Each recommendation was accompanied by illustrations from the United States Center for Disease Control and Prevention (CDC). For each recommendation, participants reported (1) how many efforts they made in the past 7 days to follow the recommendation, and (2) how many efforts they were planning to make in the upcoming 7 days to follow the recommendation, using scales from 1 (“none at all”) to 7 (“a lot”). Participants’ responses across all four recommendations were averaged into two single measures: Past Efforts and Future Efforts.¹⁰

⁹ Post-hoc power analyses using G*Power showed that the study had 99% power to detect a medium-sized effect ($f^2 = .15$) using linear multiple regression with 2 predictors.

¹⁰ Measures of past efforts ($Mean = 6.01, SD = .890; \alpha = .756$) and future efforts ($Mean = 6.34, SD = 0.980; \alpha = .799$) showed good internal consistency.

Afterward, participants were asked to rate to what extent they were described by a series of 15 statements on a scale from 1 (“definitely false”) to 6 (“definitely true”). Four of these statements (e.g. “I believe we should look after those who are vulnerable”) were adapted from Moral Foundations Questionnaire’s Care/Harm scale (Graham et al. 2011). Average ratings across these four statements constitute our measure of Moral Beliefs.¹¹ For each Moral Belief statement there was a paired Moral Emotion statement (e.g. “I feel *admiration* for people who help the weak and needy”, see Table 1 for the full list). Average ratings across these four statements constitute our Moral Emotions measure.¹² Three statements measuring participants’ trait anxiety (Spielberger 1986) and three statements measuring trait pathogen disgust (Tybur, Lieberman, and Griskevicius 2009) were included as fillers. One item (“I believe it is better to do good than to do bad”) was used as an attention check (participants who gave ratings <4 were excluded from the analyses).¹³ Statements’ order of presentation was randomized.

Moral Beliefs about Harm

Moral Emotions towards Harm

I believe we should look after those who are vulnerable.

I feel admiration for people who help the weak and needy.

I believe we should avoid people’s suffering first and foremost.

I feel guilty when people suffer as a consequence of my actions.

I think that being cruel is one of the worst

I feel outraged when someone is cruel

¹¹ This moral beliefs measure showed acceptable internal consistency ($\alpha = .59$).

¹² This moral emotions measure showed good internal consistency ($\alpha = .74$).

¹³ One might want to contend that some participants might have denied this statement because they are cynical or amoral rather than inattentive. Excluding this participants might not necessarily be bad, but this limitation should be noted. Furthermore, being reminded of the goodness of doing good might make participants more prone to behave morally. This, however, doesn’t challenge our results, as they concern the factors that explain moral behavior rather than moral behavior itself.

things one could do.

in their actions.

I think it is never right to kill a human being. Any human killing makes me feel disgusted.

Table 1. Moral Emotion and Moral Belief statements in Study 1 (emphasis added).

4.2. Results

Linear regression analyses using Moral Beliefs as the only independent variable found a significant effect of Moral Beliefs on both Past Efforts¹⁴ and Future Efforts¹⁵ to comply with official health recommendations during the COVID-19 pandemic, replicating previous results (see §3.2).

However, when both Moral Beliefs and Moral Emotions were used as independent variables, only Moral Emotions had a significant effect on participants' Past Efforts¹⁶ and Future Efforts¹⁷ to comply with official health recommendations. Moral Beliefs showed no significant effects¹⁸ (see Figure 1).

Collinearity diagnostics showed that multicollinearity was not a concern in none of our models (Tolerance was $>.20$ for all variables; Menard 2011), and thus that Moral

¹⁴ $B = .485, SE = .079, CI [.328, .642], t(169) = 6.10, p < .001.$

¹⁵ $B = .489, SE = .089, CI [.313, .665], t(169) = 5.49, p < .001.$

¹⁶ $B = .381, SE = .111, CI [.161, .601], t(169) = 3.43, p = .001.$

¹⁷ $B = .565, SE = .121, CI [.325, .804], t(169) = 4.65, p < .001.$

¹⁸ Past Efforts $B = .148, SE = .125, CI [-.098, .395], t(169) = 1.19, p = .236$; Future Efforts $B = -.010, SE = .136, CI [-.279, .259], t(169) = -.07, p = .942.$

Beliefs and Moral Emotions scores are independent.

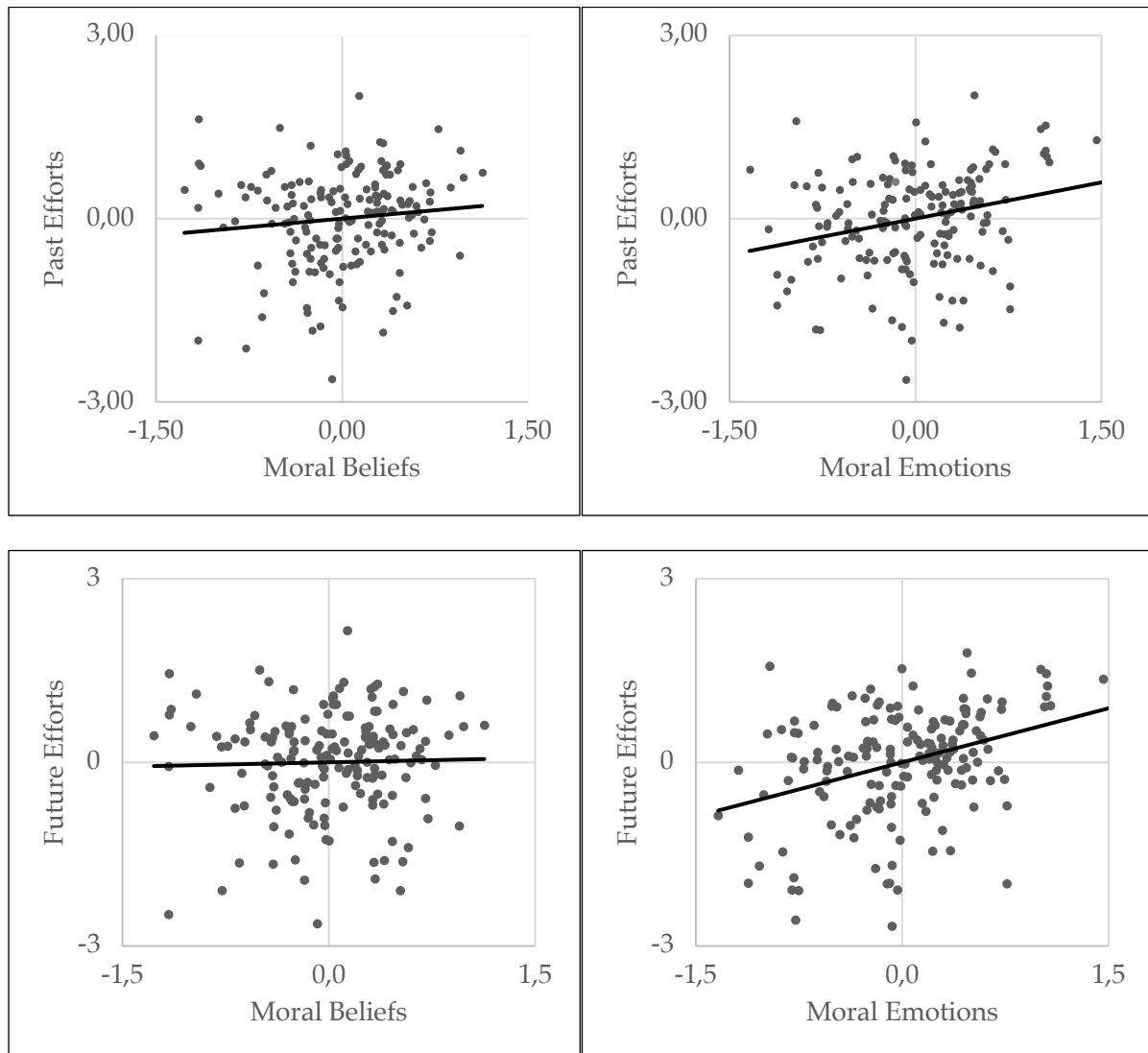


Figure 1. Partial plots for Moral Beliefs and Moral Emotions (x-axis) and Past / Future Efforts (y-axis) in Study 1.¹⁹

4.3. Discussion

The results of Study 1 suggest that the previously recorded effect of moral beliefs on moral choices can be explained by co-occurring but belief-independent moral

¹⁹ Note that the dots in partial plots show the *residuals* of the plotted variables for each participant, and not the actual values. Residuals are the negative or positive distance to the regression line in a regression where the other is used as an independent variable, and the variable at hand is used as the dependent variable. This way, they show the variance in a given variable that is not explained by the other variables. For example, the variance in Moral Emotions / Future efforts that is not explained by Moral Beliefs.

emotions. This supports Humean views of moral motivation, according to which moral beliefs themselves are motivationally inert.

Some might worry that the statements used to measure participants' moral beliefs do not concern the specific acts that they are supposed to explain, i.e., following health recommendations during the COVID-19 pandemic. Instead, they concern acts of harm/care in general. If participants believe that following health recommendations is not an instance of avoiding harm, they might believe that they should avoid harm but not believe that they should follow health recommendations. However, the same applies to our measure of moral emotions. If participants believe that following health recommendations is not an instance of avoiding harm, they might have strong emotions towards harm avoidance but not towards following health recommendations. Thus, this cannot explain why Moral Emotions explained participants' efforts to follow health recommendations better than Moral Beliefs.

There are, however, other limitations to this study. First, the considered dependent variable (health behaviors during the COVID-19 pandemic) is not prototypically moral. Second, participants' moral motivation was measured using self-reported efforts to adopt healthy behaviors. Although previous research shows no evidence of social desirability bias in these measures (Jensen, 2020), some might still worry that participants' willingness to engage in health behaviors might not translate into actual behavior. To account for these worries, the next studies will consider participants' fairness by measuring their actual fair behavior in economic games.

5. Studies 2a-b: Dictator Game

Previous research has shown that moral beliefs regarding Fairness (as measured by the MFQ) are related to moral fairness behavior in a Dictator Game (Schier, Ockenfels, and Hofmann 2016). Here, we will investigate whether these results can be explained

away in terms of moral emotions towards Fairness.

5.1. Study 2a: Dictator Game – Low stakes

5.1.1. Methods

299 US participants were recruited through Prolific and completed the survey for a monetary payment. 1 failed the attention check (see below), leaving a final sample of 298 participants (152 female, 1 non-binary, Mean age = 45.02, Age range 18-82).²⁰

Participants rated a series of statements about themselves using scales from 1 (“definitely false”) to 6 (“definitely true”). Five statements (e.g. “I believe that governments should ensure that everyone is treated fairly”) are adapted from the Moral Foundations Questionnaire’s Fairness/Reciprocity scale (Graham et al. 2011). Average ratings across these five statements constitute our Moral Beliefs measure.²¹ The other five statements (e.g. “I feel admiration for politicians that work towards preserving fairness”) are parallel to those regarding moral beliefs (see Table 2 for the full list). Average ratings across these five statements constitute our Moral Emotions measure.²² Three statements measuring participants' trait anxiety (Spielberger 1986) and three statements measuring trait pathogen disgust (Tybur, Lieberman, and Griskevicius 2009) are included as fillers. One statement (“I think is better to do good than to do bad”) is used as an attention check (participants who give ratings of <4 were excluded from the analyses). Statements’ order of presentation was randomized.

Moral Beliefs about Fairness

Moral Emotions towards Fairness

I believe that governments should ensure *I feel admiration* for politicians that

²⁰ Sensitivity analyses using G*Power show that this provides enough power to detect an effect as small as $f^2 = 0.036$ at the standard 0.05 error probability using multiple linear regression with 2 predictor variables.

²¹ This moral beliefs measure showed good internal consistency ($\alpha = .72$).

²² This moral emotions measure showed good internal consistency ($\alpha = .75$).

that everyone is treated fairly.	work toward preserving fairness.
<i>I believe</i> that we should avoid unfairness first and foremost.	<i>I feel guilty</i> when I am unfair in my actions.
<i>I think</i> that denying someone's rights is one of the worst things one could do.	<i>I feel outraged</i> when someone denies other people's rights.
<i>I think</i> justice is the most important requirement for a society.	<i>I feel disgusted</i> when it looks like there is no fairness in the world.
<i>I think</i> that we should not treat some people differently than others.	<i>I feel</i> gratitude when we are all treated in the same way.

Table 2. Moral Emotion and Moral Belief statements in Studies 2 and 3 (emphasis added).

Afterward, participants provided demographic information (Gender, Age, Ethnicity). On the last page of the survey, participants were presented with a Dictator Game, which was framed within a cover story about a raffle with a £10 price. Participants are told that, as a participant with an odd-numbered ID, they will play the role of a Decider. Deciders are given 10 raffle tickets and get to choose how to distribute those tickets between them and the participant after them (the Receiver), who will have no tickets to start with. In reality, all participants act as Deciders. Participants used a 0 to 10 slider to choose how many raffle tickets they would like to transfer to the next participant. This constitutes our Donation variable.

5.1.2. Results

Linear regression analyses using Moral Beliefs as the only independent variable found

a significant effect of Moral Beliefs on participants' donations,²³ replicating previous results (see §3.2).

When both Moral Beliefs and Moral Emotions were used as independent variables, results again showed a significant effect of participants' Moral Beliefs on the number of raffle tickets they donated.²⁴ There was no significant effect of Moral Emotions²⁵ (see Figure 2).

Collinearity diagnostics showed that multicollinearity was not a concern (Tolerance was $>.20$ for all variables; Menard 2011), and thus that Moral Beliefs and Moral Emotions scores are independent.

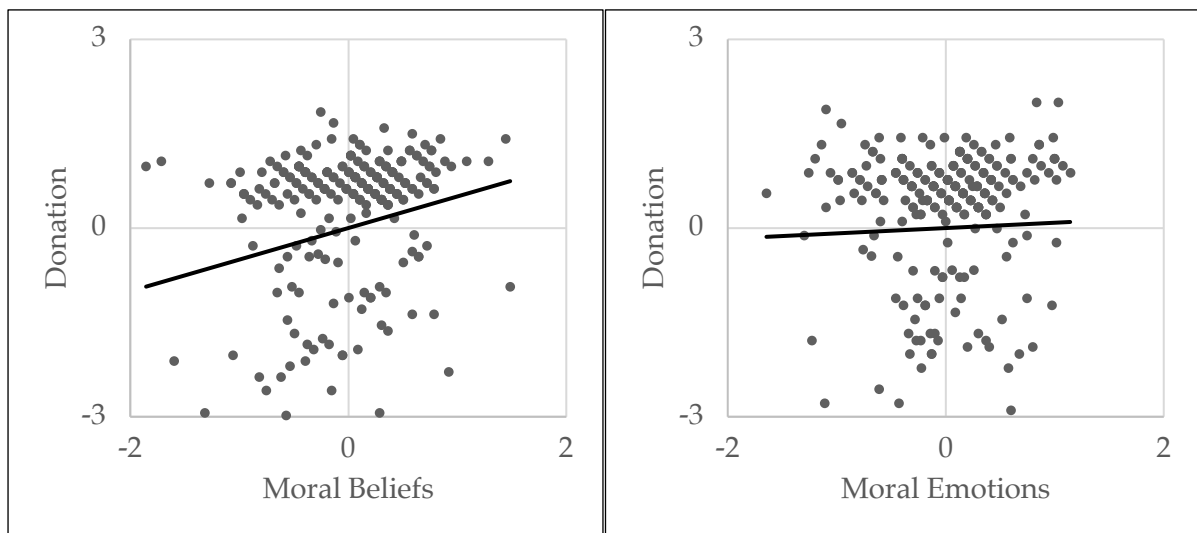


Figure 2. Partial plots for Moral Beliefs and Moral Emotions (x-axis) and Donation (y-axis) in Study 2a.

5.2. Study 2b: Dictator Game – High Stakes

Study 2b uses the same design as Study 2a, but it increases the stakes in the Dictator Game. Instead of donating tickets to enter a raffle with a £10 prize, participants' will

²³ $B = .605$, $SE = .152$, $CI [.305, .905]$, $t(297) = 3.97$, $p < .001$

²⁴ $B = .621$, $SE = .208$, $CI [.212, 1.03]$, $t(297) = 2.99$, $p = .003$

²⁵ $B = -.025$, $SE = .215$, $CI [-.448, .398]$, $t(297) = -.11$, $p = .909$

be donating tickets to enter a raffle with a £300 prize.

5.2.1. *Methods*

299 US participants were recruited through Prolific and completed the survey for a monetary payment. 5 participants failed the attention check (see below), leaving a final sample of 294 participants (205 female,²⁶ 12 non-binary, Mean age = 30.00, Age range 18-66).^{27 28}

Materials and procedure were the same as in Study 2a (see §5.1.1),²⁹ but the raffle price was £300 (instead of £10 as in Study 2a).

5.2.2. *Results*

Linear regression analyses using Moral Beliefs as the only independent variable found no significant effect of Moral Beliefs on participants' donations.³⁰

When both Moral Beliefs and Moral Emotions were used as independent variables, results showed a significant effect of participants' Moral Emotions on the number of raffle tickets they donated.³¹ There was no significant effect of Moral beliefs³² (see Figure 3).

Collinearity diagnostics showed that multicollinearity was not a concern in any of our models (Tolerance was $>.20$ for all variables; Menard 2011), and thus that Moral Beliefs

²⁶ This constitutes a high percentage of the total sample (69%). However, analyses including gender (female vs. non-female) as an independent variable did not change the results.

²⁷ Some might worry that the large percentage of female participants will affect our results. However, including sex (female vs. non-female) in our regression analysis didn't change the results, and no significant effect of sex on moral behavior was found ($p > .24$).

²⁸ Sensitivity analyses using G*Power showed that the study had enough power to detect an effect as small as $f^2 = 0.037$ at the standard 0.05 error probability using multiple linear regression with 2 predictor variables.

²⁹ Measures of moral beliefs ($\alpha = .72$) and moral emotions ($\alpha = .72$) showed good internal consistency.

³⁰ $B = .258, SE = .184, CI [-.103, .619], t(293) = 1.40, p = .161$.

³¹ $B = 1.073, SE = .268, CI [.546, 1.60], t(293) = 4.01, p < .001$.

³² $B = -.351, SE = .235, CI [-.813, .111], t(297) = -1.49, p = .136$.

and Moral Emotions scores are independent.

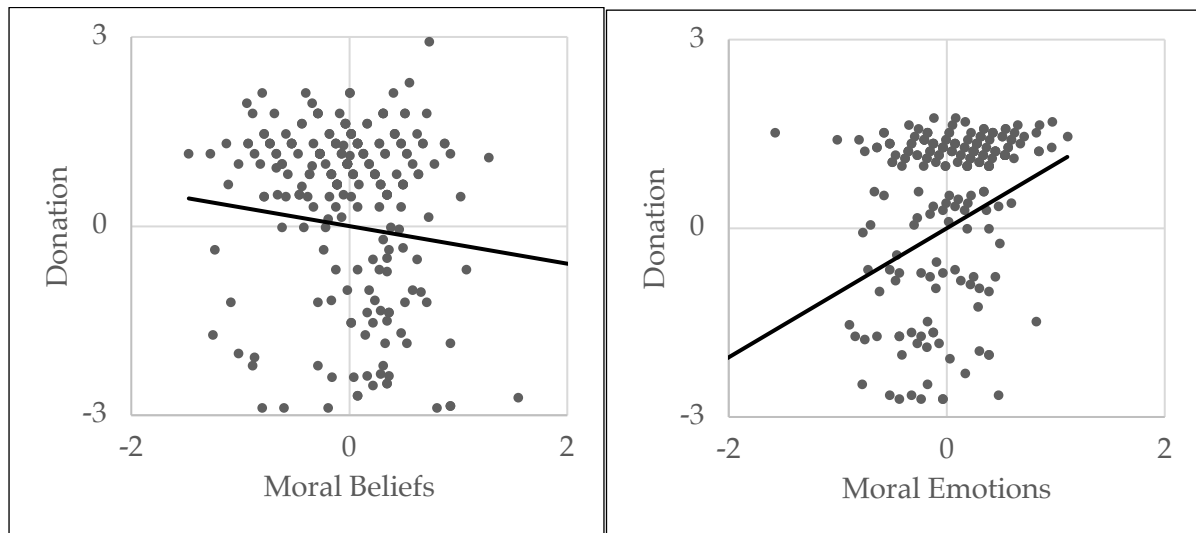


Figure 3. Partial plots for Moral Beliefs and Moral Emotions (x-axis) and Donation (y-axis) in Study 2b.

5.3. Discussion

The results of Studies 2a-b show an intriguing asymmetry. In Study 2a, moral beliefs showed a significant effect on participants' fair behavior. But in Study 2a, moral emotions did a better job at explaining participants' behavior. The only difference between the two studies is how much participants could lose by acting fairly. In Study 2a (low stakes), being fair involves lower chances to win £10. In Study 2b (high stakes), being fair involves lower chances to win £300. There are at least two possible explanations for this asymmetry.

One explanation is that low-stakes situations do not provide appropriate tests of people's moral motivation. Note that participants in Studies 2a-b reported their moral beliefs before they decided how many tickets to donate. If participants in low-stakes scenarios are indifferent to this decision, they might have acted in accordance with the moral beliefs they previously reported just to reduce the dissonance between their

reported beliefs and their actions (Festinger 1957).³³ Under this view, moral beliefs themselves do not motivate. Instead, participants in low-stake scenarios are motivated by the desire to reduce dissonance. This desire, however, would not be salient or strong enough to determine behavior in high-stakes scenarios.

Another explanation is that moral beliefs themselves are motivationally efficacious, but only when stakes are low. When there is not much to lose by doing the right thing, agents act accordingly to their moral beliefs because, absent countervailing desires or emotions, moral beliefs themselves are enough to drive action. But once something is at stake, other motivations override the influence of moral beliefs because moral beliefs have lower motivational force.

Both explanations put pressure on Anti-Humean views of moral motivation. The first explanation suggests that moral beliefs are motivationally inert. The second interpretation is compatible with Anti-Humeanism, but might reduce the motivational force of moral beliefs to the extent of having no practical relevance. This is so because countervailing motives pervade real-life situations. At any given time, desirable paths of action other than acting in accordance with our moral beliefs will be open for us. If moral beliefs motivate action only when there are no countervailing motives, then moral beliefs might play no role in determining our actions. This option preserves Motivation Anti-Humeanism, but at a very high cost.

6. Conclusion

In this paper, I showed that the use of hypothetical cases to extract conclusions regarding the (lack of) motivational power of moral beliefs faces important limitations. I argued that these limitations can be addressed using empirical research

³³ Arguably, unfair behavior is more dissonant with moral belief reports (e.g., “I believe that we should avoid unfairness first and foremost”) than moral emotion reports (e.g., “I feel guilty when I am unfair in my actions”).

tools, and presented a series of studies doing so.

The results of the studies show that, when the stakes are high, the apparent motivational force of beliefs is in fact explained by co-occurring moral emotions. This supports Humean views of moral motivation. The results regarding low-stake situations, however, are open to both Humean and “watered-down” Anti-Humean interpretations.

In moral practice, it probably won't matter if moral beliefs don't motivate us much or at all. Arguably, most real-life moral choices involve countervailing motives with more than a little motivational strength, making moral beliefs irrelevant in any case. However, the situation might be different with regards to ethical theory. Accepting that moral beliefs have some motivational force (even if very low) could be enough to solve the Moral Problem (see Introduction)³⁴ while rejecting that moral beliefs have motivational force would prompt us to reject one of the other claims involved in the puzzle. Future research should help us decide between competing interpretations of the results regarding low-stakes situations presented in this paper.

Overall, the results presented in this paper put pressure on Anti-Humean views of moral motivation, as they suggest that moral beliefs have little or no motivational force.

With regards to methodology, I showed that using empirical research tools improves upon the use of hypothetical cases of moral motivation by ruling out alternative interpretations. Note, however, that the empirical investigations presented in this paper build on extant hypothetical cases and the logical tools involved in the discussion of these cases. In this sense, the studies presented in this paper do not

³⁴ However, one might expect moral judgments to have more than a little motivational force. Thus, the claim that moral beliefs have very low motivational force might still be at odds with Judgment Internalism.

oppose, but rather continue extant work regarding cases. Hopefully, this paper paves the way for more empirical investigations, as well as discussions on the best ways to measure and test the relations between moral behavior, moral beliefs, and moral emotions.

References

- Allison, Paul D. 1999. *Multiple Regression: A Primer*. Thousand Oaks: Pine Forge Press.
- Arruda, Caroline T. 2017. "What Kind of Theory Is the Humean Theory of Motivation?" *Ratio* 30 (3): 322–42. <https://doi.org/10.1111/rati.12145>.
- Blackburn, Simon. 1998. *Ruling Passions: A Theory of Practical Reasoning*. *The Philosophical Review*. Oxford: Clarendon Press. <https://doi.org/10.2307/2693626>.
- Bromwich, Danielle. 2010. "Clearing Conceptual Space for Cognitivist Motivational Internalism Author(s):" *Philosophical Studies* 148 (3).
- Buckland, Luke, Matthew Lindauer, David Rodríguez-Arias, and Carissa Véliz. 2021. "Testing the Motivational Strength of Positive and Negative Duty Arguments Regarding Global Poverty." *Review of Philosophy and Psychology*, May, 1–19. <https://doi.org/10.1007/s13164-021-00555-4>.
- Cova, Florian, Julien Deonna, and David Sander. 2015. "Introduction: Moral Emotions." *Topoi* 34 (2): 397–400. <https://doi.org/10.1007/s11245-015-9345-0>.
- Curry, Oliver Scott, Matthew Jones Chesters, and Caspar J. van Lissa. 2019. "Mapping Morality with a Compass: Testing the Theory of 'Morality-as-Cooperation' with a New Questionnaire." *Journal of Research in Personality* 78 (February): 106–24. <https://doi.org/10.1016/J.JRP.2018.10.008>.
- Dancy, Jonathan. 1993. *Moral Reasons*. Blackwell. <https://www.wiley.com/en-us/Moral+Reasons-p-9780631187929>.
- Darwall, SL. 1983. *Impartial Reason*. Cornell University Press.

- Díaz, Rodrigo, and Florian Cova. 2022. "Reactance, Morality, and Disgust: The Relationship between Affective Dispositions and Compliance with Official Health Recommendations during the COVID-19 Pandemic." *Cognition and Emotion* 36 (1): 120–36. <https://doi.org/10.31234/osf.io/5zrqx>.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*.
- Field, Andy. 2013. *Discovering Statistics Using IBM SPSS Statistics*. London: SAGE Publications. <https://doi.org/10.1016/B978-012691360-6/50012-4>.
- Graham, Jesse, Brian A. Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H. Ditto. 2011. "Mapping the Moral Domain." *Journal of Personality and Social Psychology* 101 (2): 366–85. <https://doi.org/10.1037/a0021847>.
- Gregory, Alex. 2017. "Might Desires Be Beliefs about Normative Reasons for Action?" In *The Nature of Desire*, edited by J. Deonna and F. Lauria, 106–17. Oxford University Press.
- Haidt, Jonathan. 2003. "The Moral Emotions." In *Handbook of Affective Sciences*, edited by Richard J. Davidson, Klaus R. Scherer, and H. H. Goldsmith, 852–70. Oxford: Oxford University Press. <https://doi.org/10.1093/mnras/stx1358>.
- Hume, David. 1739. *A Treatise of Human Nature*. Edited by P. Nidditch. 1978th ed. Oxford: Oxford University Press.
- Kumar, Victor. 2016. "The Empirical Identity of Moral Judgment." *Philosophical Quarterly* 66 (265): 783–804. <https://doi.org/10.1093/pq/pqw019>.
- Lindauer, Matthew, Marcus Mayorga, Joshua Greene, Paul Slovic, Daniel Västfjäll, and Peter Singer. 2020. "Comparing the Effect of Rational and Emotional Appeals on Donation Behavior." *Judgment and Decision Making* 15 (3): 413–20. <https://doi.org/10.31234/osf.io/5wjuv>.
- Little, Margaret Olivia. 1997. "Virtue as Knowledge: Objections from the Philosophy of Mind." *Nous*. <https://doi.org/10.1111/0029-4624.00035>.

- May, Joshua. 2013. "Because I Believe It's the Right Thing to Do." *Ethical Theory and Moral Practice* 16 (4): 791–808. <https://doi.org/10.1007/s10677-012-9394-z>.
- . 2018. *Regard for Reason in the Moral Mind*. Oxford: Oxford University Press.
- Menard, Scott. 2011. *Applied Logistic Regression Analysis*. *Applied Logistic Regression Analysis*. Sage Publications. <https://doi.org/10.4135/9781412983433>.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Oxford University Press. <http://www.tandfonline.com/doi/full/10.1080/00455091.1973.10716052>.
- Nussbaum, Martha. 2001. *Upheavals of Thought*. Cambridge: Cambridge University Press. <https://doi.org/10.5840/gfpj200223213>.
- Price, Huw. 1989. "Defending Desire-as-Belief." *Mind*. Oxford Academic. <https://doi.org/10.1093/mind/XCVIII.389.119>.
- Prinz, Jesse. 2007. *The Emotional Construction of Morals*. New York: Oxford University Press.
- . 2011. "Against Empathy." *Southern Journal of Philosophy* 49 (SUPPL. 1): 214–33. <https://doi.org/10.1111/j.2041-6962.2011.00069.x>.
- . 2015. "An Empirical Case for Motivational Internalism." In *Motivational Internalism*, edited by G. Bjornsson, C. Strandberg, R. Ollinder, J. Eriksson, and F. Bjorklund, 61–84. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199367955.003.0004>.
- Prinz, Jesse, and Shaun Nichols. 2010. "Moral Emotions." In *The Moral Psychology Handbook*. Oxford: Oxford University Press. <https://doi.org/10.1016/B978-0-12-375000-6.00099-9>.
- Rosati, Connie S. 2016. "Moral Motivation." In *Stanford Encyclopedia of Philosophy*. <https://doi.org/10.5860/choice.41sup-0181>.
- Roskies, Adina. 2003. "Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy' [1]." *Philosophical Psychology* 16 (1): 51–66.

<https://doi.org/10.1080/0951508032000067743>.

Sassenberg, Kai, and Lara Ditrich. 2019. "Research in Social Psychology Changed Between 2011 and 2016: Larger Sample Sizes, More Self-Report Measures, and More Online Studies." *Advances in Methods and Practices in Psychological Science* 2 (2): 107–14. <https://doi.org/10.1177/2515245919838781>.

Schier, Uta K., Axel Ockenfels, and Wilhelm Hofmann. 2016. "Moral Values and Increasing Stakes in a Dictator Game." *Journal of Economic Psychology* 56: 107–15. <https://doi.org/10.1016/j.joep.2016.06.004>.

Schroeder, Timothy. 2004. *Three Faces of Desire*. <http://www.amazon.com/Three-Faces-Desire-Philosophy-Mind/dp/019517237X>.

Schurman, J. G. 1894. "The Consciousness of Moral Obligation." *The Philosophical Review* 3 (6): 641. <https://doi.org/10.2307/2175823>.

Schwitzgebel, Eric, Bradford Cokelet, and Peter Singer. 2020. "Do Ethics Classes Influence Student Behavior? Case Study: Teaching the Ethics of Eating Meat." *Cognition* 203 (December 2019): 104397. <https://doi.org/10.1016/j.cognition.2020.104397>.

———. 2021. "Students Eat Less Meat After Studying Meat Ethics." *Review of Philosophy and Psychology*, no. August.

Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. <https://doi.org/10.1215/00318108-2006-017>.

Sinhababu, Neil. 2009. *The Humean Theory of Motivation Reformulated and Defended*. *Philosophical Review*. Vol. 118. <https://doi.org/10.1215/00318108-2009-015>.

———. 2017. *Humean Nature*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198783893.001.0001>.

Smith, Michael. 1987. "The Humean Theory of Motivation." *Mind* 96 (381): 36–61.

———. 1994. *The Moral Problem*. Oxford: Basil Blackwell. <https://www.wiley.com/en->

us/The+Moral+Problem-p-9780631192466.

Sobel, David. 2013. "Motivation, Humean Theory Of." In *International Encyclopedia of Ethics*, 3449–55. <https://doi.org/10.1002/9781444367072.wbiee208>.

Spielberger, C. D. 1986. "State-Trait Anxiety Inventory for Adults." *APA PsycTests*. <https://psycnet.apa.org/doiLanding?doi=10.1037/t06496-000>.

Swartzler, Steven. 2013. "Appetitive Desires and the Fuss about Fit." *Philosophical Studies* 165 (3).

Tamul, Dan, Malte Elson, James Ivory, Jessica Hotter, Madison Lanier, Jordan Wolf, and Nadia Martínez-Carrillo. 2020. "Moral Foundations' Methodological Foundations: A Systematic Analysis of Reliability in Research Using the Moral Foundations Questionnaire." <https://doi.org/10.31234/osf.io/shcgvy>.

Tangney, June Price, Jeff Stuewig, and Debra J Mashek. 2007. "Moral Emotions and Moral Behavior." *Annual Review of Psychology* 58: 345–72. <https://doi.org/10.1146/annurev.psych.56.091103.070145>.

Tappolet, Christine. 2016. *Emotions, Values, and Agency*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199696512.001.0001>.

Tybur, Joshua M., Debra Lieberman, and Vladas Griskevicius. 2009. "Microbes, Mating, and Morality: Individual Differences in Three Functional Domains of Disgust." *Journal of Personality and Social Psychology* 97 (1): 103–22. <https://doi.org/10.1037/a0015474>.

Yip, Brandon. 2021. "Emotions as Modulators of Desire." *Philosophical Studies*, no. June. <https://doi.org/10.1007/s11098-021-01697-y>.

Zakharin, Michael, and Timothy C. Bates. 2021. "Remapping the Foundations of Morality: Well-Fitting Structural Model of the Moral Foundations Questionnaire." *PLoS ONE* 16 (10 October): 1–26. <https://doi.org/10.1371/journal.pone.0258910>.

Zhang, Hansong, Joshua N. Hook, and Kathryn A. Johnson. 2016. "Moral Foundations

Questionnaire." In *Encyclopedia of Personality and Individual Differences*, 1–3. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1252-1.