

# Morality justifies motivated reasoning

Corey Cusimano (cusimano@princeton.edu)

Tania Lombrozo (lombrozo@princeton.edu)

Department of Psychology, Princeton University,  
Princeton, NJ 08540, USA

## Abstract

A great deal of work argues that people demand impartial, evidence-based reasoning from others. However, recent findings show that moral values occupy a cardinal position in people's evaluation of others, raising the possibility that people sometimes prescribe morally-good but evidentially-poor beliefs. We report two studies investigating how people evaluate beliefs when these two ideals conflict and find that people regularly endorse motivated reasoning when it can be morally justified. Furthermore, we document two ways that moral considerations result in prescribed motivated reasoning. First, morality can provide an alternative justification for belief, leading people to prescribe evidentially unsupported beliefs to others. And, second, morality can affect how people evaluate the way evidence is weighed by lowering or raising the threshold of required evidence for morally good and bad beliefs, respectively. These results illuminate longstanding questions about the nature of motivated reasoning and the social regulation of belief.

**Keywords:** belief; ethics of belief; moral judgment; motivated reasoning

## Introduction

A newlywed learns that those who marry in similar circumstances to her own have a high probability of divorce. Should she believe her marriage is likely to end in divorce (consistent with the evidence), or that her marriage will succeed (consistent with her vows and her duty to her spouse)? A man learns that a lifelong friend is likely to have committed a crime, though the friend denies it. Should the man believe his friend is guilty (consistent with the evidence), or give him the benefit of the doubt (consistent with his loyalties as a friend)?

A longstanding view in epistemology states that beliefs are justified or permissible to the extent they are based on sufficient evidence, sound inference, and impartial reasoning (e.g., Clifford, 1877). On a view like this, the newlywed, the friend, and impartial observers – holding the totality of their evidence fixed – should adopt the very same beliefs. But an alternative view is that moral considerations play a legitimate role: they not only influence how beliefs are formed, or how they affect behavior, but which beliefs one *ought* to hold – either by offering non-evidential grounds for belief (e.g., Stroud, 2006) or by affecting how one ought to interpret and weigh one's evidence (e.g., Pace, 2011).

In the current paper we investigate whether people believe non-evidential considerations impact what one ought to believe. In particular, we investigate how moral and

evidential considerations interact in the beliefs we sanction for others (Study 1) and in how we judge those who have formed evidential and non-evidential beliefs (Study 2). These questions have both practical and theoretical value, as people's standards for belief predict not only how they evaluate others' beliefs, but also when they are motivated to change or update their own (e.g., Pennycook et al., 2019; Stahl, Skitka, & van Prooijen, 2014). If people do regularly prescribe partial reasoning on the basis of non-evidential consideration, this would also help explain how some motivated beliefs come about and persist.

## Prior work on lay belief prescriptions

Although people regularly fail to reason objectively themselves (Kunda, 1990), there is reason to think that they endorse objectivity as the proper basis for evaluating beliefs. For instance, even though people are often biased in their belief formation, they nevertheless believe that their own beliefs reflect an objective assessment of the evidence (e.g., Ross & Ward, 1996), update their beliefs when they think they uncover bias (Wegener, Silva, Petty, & Garcia-Marques, 2012), and try to form beliefs so that they can justify them to impartial others (Kunda, 1990). Some work suggests that people evaluate others' beliefs the same way: people prefer not to cooperate with others who exhibit partiality (e.g., Kennedy & Pronin, 2008), and many condemn those who form beliefs without sufficient evidence (Stahl et al., 2014).

However, there is also reason to think that people reject their commitment to impartiality when it conflicts with their moral values. In other domains, moral values enjoy an elevated status, outweighing and resisting comparison to non-moral considerations (Baron & Spranca, 1997; Tetlock et al., 2000). Similarly, moral virtues may displace epistemic virtues when they conflict. Consistent with this, moral virtues (e.g., loyal, just, kind) are seen as more important to judging someone's character than are epistemic virtues such as being logical or intelligent (Goodwin, Piazza, & Rozin, 2014).

At present, it is an open question whether people ever prescribe believing against the evidence when evidence and moral obligation conflict. Tenney, Logg, and Moore (2015) found that people believe that optimistic beliefs can improve performance (and therefore be self-fulfilling) and, on that basis, sometimes prescribe overly-optimistic beliefs to others (see also Armor, Masey, & Sackett, 2008). However, these authors did not independently measure what participants believed the target had evidence for, and so could not test whether participants prescribed beliefs that diverged from

that evidence. Cao, Kleiman-Weiner, and Banaji (2019) report that people condemn others' morally-charged beliefs (e.g., that a surgeon is more likely to be male than female) despite holding those beliefs for rational reasons themselves. It is possible that the belief violates a moral norm to which people are more sensitive when evaluating others than themselves. However, as the authors note, people may have condemned others because they assumed those others formed their beliefs in an evidentially poor way (e.g., baseless assumptions) rather than via Bayesian reasoning. Thus, prior work is suggestive that people may prescribe beliefs against the evidence, but this work is ultimately inconclusive.

## Two ways morality could impact belief

We will call the discrepancy between what someone ought to believe based on an objective assessment of the evidence and what someone ought to believe when taking into account moral considerations *prescribed motivated reasoning*. Work in moral philosophy and epistemology has identified two potential (non-exclusive) ways that moral concerns could lead people to prescribe motivated reasoning to others (Bolinger, 2020), described below.

The first way moral considerations could impact belief is by providing an alternative, *non-evidential justification* to believe something. On this view, moral considerations are weighed against evidential considerations when forming an "all things considered" evaluations of belief, which could result in people prescribing beliefs that are illogical or inconsistent with the evidence. In the example above, this may entail that a moral obligation to be loyal to one's friend justifies believing – against the evidence – that one's friend is blameless (e.g., Stroud, 2006). We test this prediction in Study 1. This hypothesis also predicts that people should judge motivated beliefs less harshly when the motivation is moral compared to when it is non-moral, and, likewise, judge evidence-based beliefs more harshly when they violate a moral prescription compared to when they do not. And specifically, morality should affect these judgments even after accounting for how evidentially-satisfactory the belief is. We test these predictions in Study 2.

The second way that morality could influence a belief's normative status is by altering the evidential criteria for belief (e.g., Pace, 2011). One part of belief formation is determining how to weigh evidence and what threshold of evidence to require before accepting a belief. This process varies across people and contexts (e.g., Kruglanski & Webster, 1996). For instance, the same evidence results in lower or higher confidence, and longer or shorter time before acceptance, depending on how motivated one is to avoid false beliefs (Maysel & Kruglanski, 1987). However, weighing evidence differently across situations can generate pernicious double-standards such that evidence for desired beliefs is weighed more heavily than evidence for undesired beliefs, facilitating motivated reasoning (Ditto & Lopez, 1992). Yet, it is precisely this kind of double standard – shifting the threshold for evidence – that people could morally prescribe.

This *evidential-threshold shift* hypothesis predicts that a loyal friend, relative to an objective observer with the same total information, should have a higher threshold of evidence before accepting that their friend did something wrong. Importantly, this would not be because the friend has greater prior confidence that the friend is innocent, but because the friend is subject to moral constraints that the objective observer is not. To test this possibility, in Study 1 we investigate whether, based on the same evidence, a morally-bound agent and a neutral agent can be consistent with the evidence while holding different beliefs. In Study 2, we test a related prediction that people will demand less evidence for morally good beliefs relative to morally neutral beliefs.

## Study 1

Prior work has identified several realistic situations in which moral considerations might plausibly override what one ought to believe on the basis of impartially evaluating the evidence (e.g., Basu, 2018; Bolinger, 2018; Cao et al., 2019; Stroud, 2006; Pace, 2011). Based on this work, we generated six scenarios that pit the evidence available to some believer against a moral value (see Table 1). We then tested two predictions about these cases: (1) that people will prescribe biased belief (that is, a belief that is not the most accurate), and that (2) people will shift their evidential standards (that is, they will have more stringent evidential requirements for beliefs that challenge their moral values).

## Methods

**Participants.** We recruited 839 adults (441 reported female, 395 reported male, mean age 38) from Amazon Mechanical Turk (MTurk). An additional 144 participants were excluded for failing at least one of three comprehension questions. For both studies, participation was restricted to users with US-based IP address and a 95% rating based on at least 500 HITs. Both studies were pre-registered, and approval was obtained from the Princeton IRB board.

**Methods and Design.** Participants were randomly assigned to read one of six vignettes (see Table 1). All vignettes featured situations in which the main character acquires strong but inconclusive evidence for a proposition that they have a moral reason to reject. For instance, in "Friend," the target, Adam, learns that his friend may have done something bad, but because he is loyal, he has an obligation to trust his friend's testimony. Participants were told that the main character is considering whether the proposition in question (e.g., "that John is innocent") is true, and asked to make judgments about what the character will and should believe.

*Most accurate estimate.* The first normative judgment that participants reported was what the most accurate estimate the main character could make in light of the evidence they have. To estimate what would be "objectively" most accurate, participants were told to imagine that the main character's mind was uploaded to an advanced AI that is able to "detect, catalog, and synthesize" all of the character's information and experiences. Participants then indicated what this "perfectly

Table 1: Overview of six domains used in Studies 1 and 2 where moral demands plausibly affect belief prescriptions.

Vignette	Belief	Evidence	Moral Consideration
Bully	New student will behave poorly tomorrow.	Behaved poorly on first day of class. Older siblings were poorly-behaved.	Obligation to treat every student as having high potential.
Cancer	Husband will survive his cancer.	Studies show that only 15% of those diagnosed live past 1 year.	Optimism will improve well-being for husband and family.
Marriage	Main character and wife will eventually divorce.	Reads a study showing that 70% of similar marriages end in divorce.	Vowed life-long commitment. Optimism facilitates a better relationship.
Friend	Friend is not guilty of possessing cocaine.	Drugs found in dorm room, rumors of friend associating with drug dealers.	Friend requests benefit of the doubt, loyalty to friend demands trust.
Race	Approaching Black man is dangerous.	In this neighborhood, 80% of young Black men are in a dangerous gang.	Respecting others demands you do not judge based on group statistics.
Sex	Approaching woman is surgeon rather than nurse.	In this particular dental practice, 90% of surgeons are male.	Respecting others demands you do not judge based on group statistics.

detached observer” would estimate based on that information. Participants responded to this question, and the questions below, using a slider anchored at 0% and 100% with 5% intervals.

*Evidence-based bounds on belief.* Next, participants reported the most optimistic and most pessimistic estimates that could be considered “consistent with” and “based on” the evidence available. Participants were reminded of the “most accurate” estimate they had just provided and told that the advanced AI can also “calculate the most optimistic and the most pessimistic estimates licensed by the available evidence.” Participants were instructed to report what these estimates would be “if the advanced AI made sure its estimates are based on, and consistent with, all the information that it has from [character’s] brain.”

A 0%-100% range input appeared below containing three slider handles (see Figure 1). One handle was immovable and set to the estimate they had indicated to be most accurate. Up to 15%-points to the left and right were two handles that represented evidence-bound pessimism and optimism. Participants could move these handles in five-point increments from the anchors (0% or 100%) to the “most accurate” value handle. After participants submitted their judgments for the AI, instructions appeared below directing participant to now report what the evidence-based bounds

would be for the main character. We used the deviation between participants’ estimates for the character and the AI as a measure of the “evidential threshold shift” hypothesis.

*Ought-to-believe.* On the next page, participants reported what the agent ought to believe. The default slider value was set to the “most accurate” estimate that the participant had previously provided. Deviation from this anchor would provide evidence for prescribed motivated reasoning. Furthermore, if participants provided “ought” judgments that fell outside the evidential bounds they just provided, this would provide evidence for the “alternative justification” hypothesis, which predicts that people will sometimes prescribe beliefs that are not justified on evidential grounds.

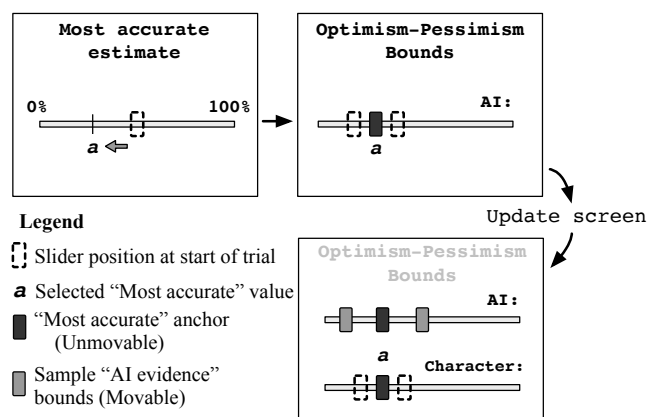
*Moral concern.* On the next screen, participants reported their agreement with a series of statements about the moral value of the main character’s beliefs in the vignette. Two statements measured commitment to the moral norm we hypothesized would be most operative in the vignette (e.g., “All else being equal, it is morally good to give your friend the benefit of the doubt”). Participants reported their agreement on a 7-point scale (anchored at “strongly disagree” and “strongly agree”).

## Results

As expected, participants’ ought estimates ( $M = 41, SD = 28$ ) were significantly more optimistic than their judgments of what is most accurate ( $M = 31, SD = 23$ ),  $F(1, 833) = 145.47, p < .001$ , with 53% of participants giving ought estimates higher than accurate estimates. The strength of this difference varied by scenario,  $F(5, 833) = 11.08, p < .001$  (Figure 2A). This offers strong support for the proposal that at least some participants prescribe motivated reasoning to others.

To test whether prescribed motivated reasoning was driven by moral concern, we averaged the two moral concern items. As expected, moral concern was associated with the extent to which ought judgments deviated from what would be most accurate to believe,  $F(1, 827) = 32.34, p < .001$ .

Participants also reported that the main characters in the scenarios were licensed to evaluate the evidence differently than the perfectly impartial, but equally informed, observer, consistent with an evidential-threshold shift. On average,



would be for the main character. We used the deviation

Figure 1. Procedure for measuring evidential threshold shifts.

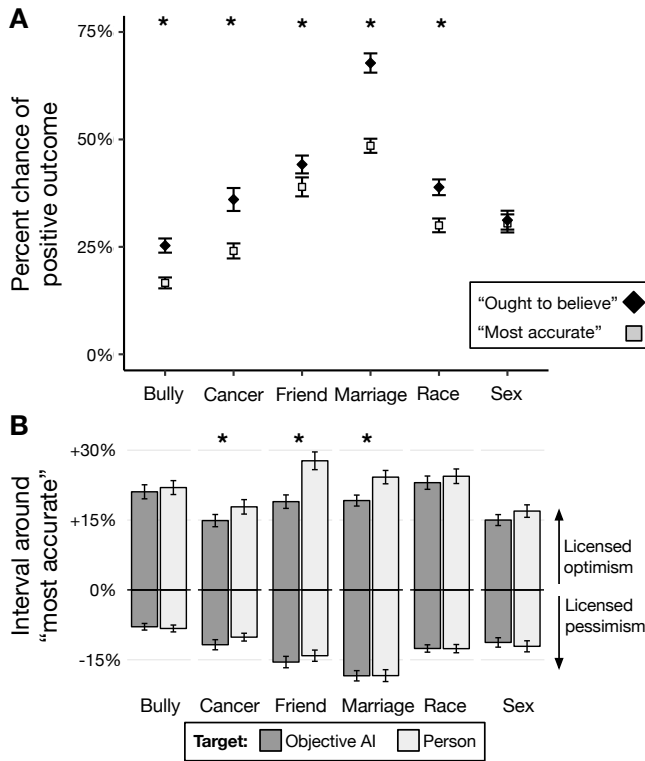


Figure 2. Main results from Study 1 across vignette. (A) Mean (and standard error) ratings for prescribed belief and most accurate judgments in Study 1. (B) Mean (and standard error) judgments of evidence-licensed optimism and pessimism for the advanced AI (dark bars) and person (light bars). \*  $p < .05$ .

participants believed the character was licensed to be more optimistic ( $M = 54$ ,  $SD = 25$ ) than the objective AI ( $M = 50$ ,  $SD = 26$ ),  $F(1, 833) = 44.29$ ,  $p < .001$ , with 42% of participants giving higher optimism bound estimates for the character than the AI. The strength of this difference varied across scenario,  $F(5, 833) = 5.44$ ,  $p < .001$  (Figure 2B). It is possible that participants believed characters were licensed to be less confident in general, rather than licensed to evaluate the evidence directionally. However, this is ruled out by the fact that there was no corresponding difference between how pessimistic the character ( $M = 19$ ,  $SD = 20$ ) and AI ( $M = 19$ ,  $SD = 20$ ) could be,  $F(1, 833) = 0.72$ ,  $p = .397$ .

We observed mixed support for an association between agent-AI differences in prescribed optimism (“prescribed optimism inaccuracy”) and moral concern. Prescribed optimism inaccuracy was greater when prescribed inaccuracy was greater,  $F(1, 827) = 13.66$ ,  $p < .001$ . However, there was no association between moral concern and prescribed optimism inaccuracy,  $F(1, 827) = 0.04$ ,  $p = .948$ .

Finally, we observed some evidence for the alternative justification hypothesis. Of the 53% of participants who prescribed more optimistic beliefs than what would be most accurate, 32% of these individuals (about 17% of the full sample) prescribed optimistic beliefs that fell outside the range they reported would be licensed by the evidence. In the

Cancer and Marriage vignettes, nearly half of participants who prescribed motivated reasoning did so by prescribing beliefs that were – by their own lights – unsupported by the evidence (44% and 50%, respectively).

## Discussion

Study 1 found strong support for morally-prescribed motivated reasoning. Participants routinely indicated that another person ought to hold an inaccurate belief, with the degree of inaccuracy positively related to the moral benefit of holding that belief. Study 1 also found that evidence for both the “non-evidential justification” and the “evidential-threshold shift” hypotheses regarding how moral values can result in prescribed motivated reasoning. In support of the evidential-threshold shift hypothesis, we observed that, in three out of five scenarios in which participants prescribed inaccuracy, they also reported that the agent had a more optimistic range of beliefs consistent with their evidence than was licensed to an equally-informed, but completely impartial observer. Lastly, in support of the alternative justification hypothesis, we observed that in some scenarios, a large proportion of participants prescribed beliefs that they themselves considered inconsistent with the evidence available to the believer. Notably, these findings replicated in the Friend scenario, where the belief in question concerned an event that occurred in the past. This wards off an alternative explanation that optimism was justified only because it was self-fulfilling (c.f. Tenney et al., 2015).

## Study 2

Study 2 builds on Study 1 by investigating people’s evaluations of beliefs that have already been formed. If people prescribe motivated reasoning to others, then they should judge a person with a moral reason to be optimistic more favorably for being overly optimistic compared to someone with the same evidence but who lacks a moral justification. The evidential-threshold shift hypothesis predicts that a person with a moral reason for optimism (relative to a non-morally justified observer) should require less evidence for adopting the optimistic belief, but more evidence for adopting the evidence-based belief. And finally, the alternative justification hypothesis predicts that people’s overall evaluations of a belief’s quality will be partially predicted by the moral quality of the belief even after accounting for the evidential quality of the belief. To test these predictions, we manipulated whether the believer had a moral or non-moral reason to be optimistic as well as whether they adopted the evidence-based or optimistic belief, and analyzed participants’ judgments across these conditions.

## Methods

**Participants.** We recruited 1,021 adults (638 reported female, 524 reported male, mean age 40) from MTurk. An additional 145 participants were excluded for failing at least one of three comprehension checks.

**Design and Methods.** Study 2 used a 2 (Reason) x 2 (Belief) x 4 (Vignette) mixed design. Within each scenario, participants read about the character in Study 1, who has a moral reason to favor the optimistic belief, and about another person with the same evidence, but without a moral reason to favor the optimistic belief (Reason manipulation). Participants learned either that these two people adopt the evidence-based belief or the optimistic belief (Belief manipulation). Finally, participants read one of four vignettes from Study 1 that readily enabled our Reason manipulation: Bully, Cancer, Friend, or Marriage.

Vignettes were minimally modified from Study 1 so that the main character discusses what they should believe with someone who, over the course of the conversation, ends up with the same relevant information. For instance, in the Friend vignette, the other observer knows the suspect well and only recently stopped being close with him for reasons unrelated to that person's character. To reinforce the idea that the two believers have the same information, participants reported what an "objective observer" would estimate based on their shared information.

Participants were then told that the two characters either adopt the evidence-based belief (e.g., that the friend is guilty), or the non-evidence-based belief (e.g., that the friend is innocent). The main character does so either based on, or after setting aside, their moral reason to be optimistic (Moral reason condition). By contrast, the new character does so based on, or after setting aside, "a strong preference" to be optimistic (Non-moral reason condition).

Participants responded to eight questions about each character's belief. Two measured "overall belief quality," including whether the belief was justified and whether it was permissible for the target to hold. Two questions measured the "moral quality" of the belief, including whether the belief was loyal (or helpful) and whether the person's reasoning was morally good. Two questions measured the "evidentiary quality" of the belief, including whether the person had sufficient evidence for their belief, and whether they should have obtained more evidence before coming to their belief. Finally, two questions measured the moral quality of the character, including whether they were trustworthy and whether they were a good person. These questions were presented in a random order.

Participants then read that the belief the two characters adopted turned out to be true (e.g., the friend was in fact innocent). For each character, participants reported their agreement (1: strongly disagree, 7: strongly agree) with statements claiming that that characters "knew" the outcome.

## Results

We computed overall belief quality, moral evaluations, epistemic evaluation, and character judgments by averaging together participants' responses to the two items for each DV separately for judgments of the morally-motivated and the non-morally-motivated characters. These judgments, as well as participants' knowledge attributions, were then submitted to separate 2 (Reason: Moral vs Non-moral) x 2 (Belief:

Evidentiary vs Optimistic) x 4 (Vignette) ANOVAs. Key results are displayed in Figure 3.

*Moral evaluations.* The moral quality of the belief depended both on whether the optimistic vs. evidence-based belief was adopted, and on whether the character had a moral reason to be optimistic,  $F(1, 1013) = 273.00, p < .001$ . For the evidence-based belief, the character who set aside their moral obligation was judged to have a morally worse belief ( $M = 3.87, SD = 1.51$ ) than the character who set aside their preference ( $M = 4.15, SD = 1.32$ ),  $F(1, 512) = 31.66, p < .001$ . But the character who adopted the optimistic belief for a moral reason ( $M = 5.75, SD = 1.13$ ) was judged to have a morally better belief than the character who adopted the same belief for a non-moral reason ( $M = 4.87, SD = 1.37$ ),  $F(1, 501) = 313.14, p < .001$ .

Similarly, participants' judgments of the moral quality of the believer's character depended on what belief they adopted and the presence of a moral justification,  $F(1, 1013) = 80.34, p < .001$ . On average, characters who adopted the biased, optimistic belief without a moral justification were judged as having worse character ( $M = 5.34, SD = 1.14$ ) than characters who adopted the optimistic belief for a moral reason ( $M = 5.85, SD = 1.02$ ),  $F(1, 501) = 145.18, p < .001$ . However, there was no overall difference in perceived character quality when the characters adopted the evidence-based belief despite moral ( $M = 4.91, SD = 1.43$ ) or non-moral ( $M = 4.97, SD = 1.22$ ) reasons,  $F(1, 512) = 1.64, p = .202$ .

*Prescribed motivated reasoning.* Consistent with our predictions, the pattern of judgments for the belief's moral quality replicated for judgments of the overall quality of the belief (Figure 3B). When the two characters adopted the evidence-based belief, the character with the moral reason to be optimistic was seen as less justified and less permitted to hold the belief ( $M = 5.04, SD = 1.62$ ) compared to the other, more socially distant character ( $M = 5.30, SD = 1.52$ ),  $F(1,$



Figure 3. Study 2 means (and standard errors) for sufficient evidence, overall quality of belief, and moral quality of belief ratings. \*  $p < .05$ .

512) = 23.00,  $p < .001$ . However, when they both adopted the optimistic belief, the character with a moral justification was considered more justified / permitted to do so ( $M = 5.02$ ,  $SD = 1.43$ ) compared to the character without moral justification ( $M = 4.34$ ,  $SD = 1.51$ ),  $F(1, 501) = 167.03$ ,  $p < .001$ .

*Alternative justification hypothesis.* We next regressed overall belief quality ratings on evidential-quality and moral-quality ratings separately for each belief condition (optimistic and evidence-based) and separately for each vignette. Evidentiary quality predicted overall belief quality in both the evidence-based belief conditions ( $bs > 0.35$ ,  $ts > 6.80$ ,  $ps < .001$ ) and the optimistic belief conditions ( $bs > 0.33$ ,  $ts > 5.12$ ,  $ps < .001$ ) across all four vignettes. In support of the alternative justification hypothesis, we observed that, even when accounting for differences in evidentiary quality, the moral quality of the belief independently predicted overall judgments in the evidence-based belief conditions ( $bs > 0.20$ ,  $ts > 3.42$ ,  $ps < .001$ ) and in the optimistic belief conditions ( $bs > 0.33$ ,  $ts > 2.65$ ,  $ps < .008$ ) across all four vignettes.

*Evidential-threshold shift.* Finally, we observed the predicted Reason x Belief interaction on participants' judgments about the evidentiary quality of the belief ( $ps < .001$ ), and on attributions of knowledge,  $F(1, 1013) = 50.07$ ,  $p < .001$ . When the characters adopted the evidence-based belief, there was no difference in knowledge attributed to the morally-motivated character ( $M = 3.87$ ,  $SD = 1.51$ ) and the non-morally-motivated character ( $M = 4.15$ ,  $SD = 1.32$ ),  $F(1, 512) = 2.67$ ,  $p = .103$ . But, when they were optimistic, the morally-motivated true belief was more often treated as knowledge ( $M = 5.75$ ,  $SD = 1.13$ ) than the non-morally motivated true belief ( $M = 4.87$ ,  $SD = 1.37$ ),  $F(1, 501) = 111.64$ ,  $p < .001$ . Within each Belief condition across all vignettes, participants' ratings of the moral quality of the belief significantly positively correlated with their ratings of the evidential quality of the belief ( $ps < .001$ ).

When we analyzed each of the judgments above within vignette, the interaction between Belief and Reason replicated in all vignettes except Bully ( $ps < .001$ ).

## Discussion

Mirroring belief prescriptions in Study 1, Study 2 showed that a moral reason to be optimistic increased the perceived permissibility and justifiability of overly-optimistic beliefs and lowered the permissibility and justifiability of evidence-based beliefs, relative to a non-moral reason. This provides converging evidence that people incorporate morality into their overall evaluations of others' beliefs, and so will sometimes positively evaluate motivated reasoning in others.

We also observed evidence for both the evidential-shift and alternative justification hypotheses. Mirroring the finding from Study 1 that the morally-justified characters were licensed to evaluate the evidence more optimistically, Study 2 found that optimistic beliefs were seen as more evidentially supported for the characters with moral reasons than those without moral reasons. Likewise, having a moral reason not to adopt an evidence-based belief, on average, increased the evidence required relative to an objective observer. In support

of the alternative justification hypothesis, we observed that the moral quality of the belief reliably predicted the overall quality of the belief even after accounting for how evidentially well-supported participants rated the belief to be. That is, the moral quality of the belief independently predicted how justified and permissible participants judged the optimistic and evidentially-supported beliefs to be.

The method we used to manipulate the presence of a moral reason meant that the morally-justified character was always closer to the person they were judging than the non-morally justified observer. Indeed, social closeness was the source of their moral justification. This raises a worry that, despite our efforts to match the evidence available to the two characters, the morally-justified character was seen as always forming beliefs based on more evidence. A related worry is that closeness creates a non-directional reasoning goal, such as being absolutely certain that what one believes is true before adopting any belief (e.g., fear of invalidity; Mayseless & Kruglanski, 1987). These alternative hypotheses predict that the socially-closer character would be judged as (i) better (because they had more information) or (ii) worse (because they needed more evidence) than the non-morally justified observer, whether they adopted the evidentiary or optimistic belief. However, the reason-by-belief interaction we observe rules these alternatives out. It was not the case that the morally-justified character was uniformly judged as more (or less) justified. Rather, the morally-justified character only received higher ratings for sufficient evidence and justification when their belief was morally good, and not when their belief was morally bad.

## General Discussion

Across two studies we show that moral concerns affect how people evaluate others' beliefs. In Study 1, people indicated that others ought to adopt beliefs that would be inaccurate relative to what a perfectly objective perceiver would believe, and that, in some cases, the morally concerned character had a wider, more optimistic set of beliefs consistent with her evidence. In Study 2, moral reasons lowered the evidentiary requirements for a belief, and people who adopted morally-good beliefs were attributed more knowledge and were seen as more justified than people who adopted equally biased beliefs for non-moral reasons. Thus, morality appears to provide an independent justification for belief and affects beliefs' evidentiary quality.

These results can potentially explain the presence and persistence of certain motivated beliefs. In particular, morally-motivated beliefs could persist in part because people do not demand that they or others reason accurately or acquire equal evidence for their beliefs (Metz, Weisburg, & Weisburg, 2018). These findings also invite a re-interpretation of some classic biases, which are in general interpreted as unintentional errors (Kunda, 1990). We suggest instead that some apparent errors reflect convictions that one ought to be biased or discount evidence. Future work investigating biased belief formation should incorporate the perceived moral value of the belief.

## Acknowledgements

We thank Sara Aronowitz, Renee Bolinger, Andrew Chignell, Liz Harman, and members of the Concepts & Cognition Lab as well as members of the University Center for Human Values for valuable feedback on this research.

## References

- Armor, D. A., Massey, C., & Sackett, A. M. (2008). Prescribed optimism: Is it right to be wrong about the future? *Psychological Science, 19*, 329-331
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*, 1-16.
- Basu, R., (2018). The wrongs of racist beliefs. *Philosophical Studies, 1-19*.
- Bolinger, R. J. (2018). The rational impermissibility of accepting (some) racial generalizations. *Synthese, 1-17*.
- Bolinger, R. J. (2020) The varieties of moral encroachment. *Philosophical Perspectives*. doi:10.1111/phpe.12124
- Cao, J., Kleiman-Weiner, M., & Banaji, M.R. (2019). People make the same bayesian judgment they criticize in others. *Psychological Science, 30*, 20-31.
- Clifford, W. K. (1877). The ethics of belief. In T. Madigan (Ed.), *The ethics of belief and other essays*. Amherst, MA: Prometheus.
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of personality and social psychology, 63*, 568-584.
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*, 148-168.
- Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin, 34*, 833-848.
- Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "seizing" and "freezing". *Psychological review, 103*, 263-283.
- Mayseless, O., & Kruglanski, A. W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. *Organizational Behavior and Human Decision Processes, 39*, 162-183.
- McManus, R. M., Kleiman-Weiner, M., & Young, L. (in press). What we owe to family: The impact of special obligations on moral judgment. *Psychological Science*.
- Metz, S. E., Weisberg, D. S., & Weisberg, M. (2018). Non-scientific criteria for belief sustain counter-scientific beliefs. *Cognitive Science, 42*, 1477-1503.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480-498.
- Pace, M. (2011). The epistemic value of moral considerations: Justification, moral encroachment, and James "Will To Believe". *Noûs, 45*, 239-268.
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2019) On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Preprint*.
- Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *The Jean Piaget symposium series. Values and knowledge*. US: Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Ståhl, T., Zaal, M. P., & Skitka, L. J. (2016). Moralized rationality: Relying on logic and evidence in the formation and evaluation of belief can be seen as a moral issue. *PLOS ONE, 11*, e0166332.
- Stroud, S. (2006). Epistemic partiality in friendship. *Ethics, 116*, 498-524.
- Tenney, E.R., Logg, J.M., & Moore, D.A. (2015) (Too) optimistic about optimism: The belief that optimism improves performance. *Journal of Personality and Social Psychology, 108*, 377-399
- Tetlock, P. E., Kristel, O. V., Beth, S., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*, 853-870.
- Wegener, D. T., Silva, P., Petty, R. E., & Garcia-Marques, T. (2012). The metacognition of bias regulation. In P. Brinöl & K. DeMarree (Eds.), *Social Metacognition*. New York: Psychology Press.