Abstract

Judging others' moral character is a vital and necessary task for navigating the social world. Extending work on the function of emotions, we propose that people use others' emotions to infer their moral character. We focused on anger, an emotion that is often viewed as undesirable. We hypothesized that anger could serve as a signal of moral character, specifically when experienced after observing a third-party moral violation (i.e., when one individual behaves immorally toward another). We first examined this hypothesis by showing that people not only judged the observer who *felt* angry to be a better person (Studies 1-2), but they also trusted the observer more (Study 3). In Study 4, we found that such inferences can be drawn when anger was *displayed*, and this effect was much more pronounced for third-party violations compared to when people were treated immorally themselves. Further, we explored whether the positive effect from anger is unique from sympathy (Study 5) and cognitive recognition of the violation (6a, and 6b), and found that anger elicited a similar level of positive moral character judgment as sympathy and cognitive recognition of the violation. However, different from recognition of the violation, anger is associated with a higher expectation of behavioral engagement. These studies not only demonstrate the moral character signaling function of emotions but also contribute to an understanding of the processes by which individuals infer moral character in others.

Keywords: moral character, emotion, anger, trust, person perception

> "ANYBODY can become angry, that is easy; but to be angry with the right person, and to the right degree, and at the right time, and for the right purpose, and in the right way, that is not within everybody's power, that is not easy."
>
> -Aristotle, *The Art of Rhetoric*

Anger has a bad reputation. Across many contexts, anger is treated as an undesirable emotional response. People who are quick to anger are disparagingly referred to as "hotheaded," and many spend a great deal of money and time attending anger management classes. In dating contexts, proneness to anger is frequently considered a dealbreaker. This bad reputation is not unreasonable—anger has been linked to aggression and violence (e.g., Averill, 1983; Bohnert et al., 2003; Deffenbacher et al., 2003; Norlander, & Eckhardt, 2005; Roberton et al., 2012, 2015; Scarpa, & Raine, 2000) and to maladaptive emotional processing that can result in dysfunctional emotion regulation (Fernandez, & Kerns, 2008; Larsson et al., 2023).

However, as Aristotle pointed out over 2,000 years ago, there is a kind of anger that can be beneficial. It is on one potential benefit that we focus here—that anger, when experienced at the right time, and for the right reasons, might communicate a positive moral character. Anger is often experienced as a response to the perception that a moral violation—such as harm or injustice—has occurred (Batson et al., 2007; Hoffman, 2000; Montada & Schneider, 1989). In addition, some theories suggest that although this specific type of anger drives costly and punitive behaviors, such negative behaviors may have reputational benefits for those who engage in them (Jordan & Rand, 2020; Kurzban et al., 2007).

While a great deal of previous research has looked specifically at the behavioral consequences of anger (e.g., third-party punishment), little work has focused on the potential

reputational benefits of simply *feeling anger* in response to a moral violation. It is possible that anger not only serves the function of motivating punitive behaviors, but also serves the function of signaling an agent's moral character (Anderson et al., 2021). Given that there may be norms discouraging the sorts of punitive behaviors motivated by anger, it may be that simply experiencing or expressing the emotion—without then acting in any kind of retributive or punitive manner—may be sufficient to serve as a sincere signal of one's underlying moral character.

While previous research has demonstrated that anger is the most commonly experienced emotion in the face of unfairness (Mikula et al., 1998), it is not clear how experiencing such anger influences moral character inferences. Some past work has suggested that others may view an angry person positively and linked anger expressions with positive outcomes (Hess, 2014; Hareli et al., 2009; Tiedens, 2001). However, this research often focused on competence relevant outcomes, such as status, promotions, and compensations. These associations between competence relevant positive outcomes and anger could be a result of inferences about the person's dominance or power (Keating, 1985; Tiedens et al., 2000), making it possible that observers give angry people more credit because they are intimidated by them. Alternatively, anger signals the high likelihood that the person is going to act on that emotion and therefore act aggressively (Berkowitz & Harmon-Jones, 2004; Scherer, 1987), rather than actually thinking the other person is a good person or a likable person. To investigate whether anger can be a positive signal for a person's character, in this work, we focused on examining the value of experiencing anger on moral character inferences. Given that anger maintains a bad reputation as a negative and potentially harmful emotion, it may be that those who experience (let alone express) anger are simply judged negatively. However, because anger can be a particularly

appropriate response expressed at immoral acts (e.g., unfairness), people might infer that those who feel angry have the right kind of moral character. We test this possibility in the studies that follow, by examining evaluations of individuals who are described as experiencing anger in response to observing a moral violation occur to a third party. Specifically, we hypothesize that anger in those contexts will lead to more favorable evaluations regarding the moral character of an angry observer, functioning as one of many potential sources of character information.

Judgments of Moral Character

A growing body of research has highlighted the importance of *character* in moral judgments; people do not just evaluate the morality of particular *actions* but also evaluate the character of the *agents* involved in those actions (for reviews, see Hartman et al., 2022; Helzer & Critcher, 2018; Pizarro & Tannenbaum, 2012; Uhlmann et al., 2015). Such judgments of moral character appear to be important components of social cognition more broadly. For example, research has shown that people prioritize moral traits over other traits when judging the general positivity of a person (Goodwin, Piazza, & Rozin, 2014). Likewise, people prioritize their ingroup's moral traits over other traits like competence or likability (Leach et al., 2007). People also define personal identity largely in moral character terms (Heiphetz et al., 2018; Strohminger & Nichols, 2014). Furthermore, judgments of a person's morality more strongly predict liking for that person than do judgments of that person's competence and sociability (Hartley et al., 2016). In short, assessments of moral character play a central role in our understanding of each other.

When evaluating moral character, people are attempting to understand the agent's "moral-cognitive machinery" (Critcher et al., 2020; Helzer & Critcher, 2018)—the set of underlying psychological mechanisms that govern how that agent behaves in moral situations.

This can include information about an agent's goals, motivations, wants, meta-desires, beliefs, and other mental states (Ames & Johar, 2009; Critcher et al., 2013; Fedotova et al., 2011; Gray et al., 2012; Pizarro et al., 2003). For instance, given an agent's behavior (e.g., donating to charity), observers attempt to uncover the causes of such behavior (e.g., empathic concern for others) to predict whether such behavior may occur in the future.

One important source of information about character comes from the emotions that an individual experiences and expresses across various situations. Emotions can serve as potential sources of information about an individual's intentions and desires in a given situation (Higgins, 1998). For instance, while displays of positive affect when engaging in an action might indicate that the agent is claiming ownership or responsibility over their behavior (e.g., Tracy & Robins, 2008; Weiner, 1985), displays of negative affect while engaging in an action might indicate that the agent is distancing themselves or repudiating their action (e.g., Gold & Weiner, 2000). Accordingly, observers judge agents more positively when their prosocial behavior is accompanied by a positive emotional display (e.g., smiling) than when the same behavior is accompanied by a negative emotional display (e.g., grimacing; Ames & Johar, 2009). But some emotional responses may serve as more direct signals of moral character (rather than indirect signals of intentions or desires that might feed into judgments of character). There has been at least one set of studies demonstrating this sort of character signaling function for emotions: Anderson et al. (2021) found that observers treated an agent's experience of guilt in response to an unforeseen accident as a positive sign of their moral character. It is possible that anger can also provide a similarly positive signal of moral character in some situations. However, much of the literature on anger has focused on demonstrating that anger may do the exact opposite.

Anger and Reputation

Anger is one of the most commonly experienced emotions in daily life (Averill, 1983; Kish-Gephart, Detert, Treviño, & Edmondson, 2009; Lerner & Tiedens, 2006) and has been typically regarded as negative and undesirable (Averill 1982; Berkowitz 1990; Izard 1991; Russell 1991, Waldman, Balthazard, & Peterson, 2011). It is linked to strong action tendencies, especially with an array of antisocial or destructive responses such as hostility, aggression, punishment, and exclusion (e.g., Averill, 1982; Archer, 2000; Archer & Coyne, 2005; Berkowitz, 1990, 1993; Bushman, 2002; Berkowitz, 1993; Martin et al., 1999). In short, anger is frequently seen within psychology as causing physical and psychological harm to others.

These negative consequences are not uncommon in daily life. Anger can be the cause and the consequence of unpleasant situations. Conflicts in the workplace can arise due to anger (Gibson & Callister, 2010). Anger can also hinder interpersonal communication and negotiation (Friedman et al., 2014). In intimate relationships, anger is significantly associated with discontent and even violence (Birkley & Eckhardt, 2015). Anger is not just associated with negative events, people who get angry are generally not liked by others. For example, people infer negative traits, such as high dominance and low affiliation, from those who display angry expressions (Knutson, 1996; Karasawa, 2001; Tiedens, 2003).

Even when anger can sometimes help to procure desired results (e.g., getting a bigger concession from a negotiating partner), it is often only through invoking negative feelings in the partner (e.g., fear; Sinaceur and Tiedens, 2006, Sinaceur et al., 2011; Van Kleef et al., 2004). In addition, such benefits are often short-lived and can leave negative impressions on the partner and hurt interpersonal relationships in the long run (Axelrod, 1984, Clark et al., 1996, Van Kleef et al., 2009; Van Kleef & De Dreu, 2010). Given what we know about anger, it seems as if individuals would, in general, treat experiences and expressions of anger as undesirable.

There does appear to be an exception to this, however: moral anger (i.e., the anger that results from witnessing a moral violation that impacts someone) has been shown to be associated with a more complex pattern of behavioral consequences. For example, experiencing moral anger predicts not only aggressive behaviors but also benevolent prosocial behaviors to address the perceived harm (Lotz et al., 2011; Montada & Schneider, 1989; Van Doorn et al., 2018; Vyver & Abrams, 2014).

Some theories have proposed that moral anger is a moral emotion that motivates people to restore justice, although the means could be either antagonistic (e.g., punishing) or prosocial (e.g., helping; Lotz et al., 2011; Lyer et al., 2007; Van Doorn et al., 2014; Wakslak et al., 2007). These theories seem to share two assumptions relevant to this discussion: First, in order to feel angry in the first place, a person has to recognize that a moral violation has occurred. Second, that anger motivates an individual to take action as a result of the moral violation. That is, moral anger can be a *precursor* to retributive and restorative behaviors. If this is the case, moral anger would seem to be a good indicator that a person is likely a good person because they have the ability to discern right from wrong *and* they are motivated to act to rectify whatever violation has occurred.

Is this how people, in fact, view those who express moral anger? There is some work suggesting that this might be the case. Jordan and colleagues (2016) found that certain negative behaviors associated with anger, such as third-party punishment, can serve as a signal of the agent's trustworthiness. However, that work focuses on determining the motivation behind the costly behaviors invoked by anger (e.g., costly third-party punishment) rather than anger itself. How would people judge the person who *felt angry* at a third party's moral violation compared to those who didn't?

Present Work

Across seven studies, using a variety of experimental designs and methodologies, we examined the signaling role of moral anger for judgments of moral character. Across all studies, participants were presented with a description of an observer who witnessed a moral violation and experienced anger (or not) in response to that violation. They were then asked to make evaluations of the observer's moral character. We hypothesized that participants would judge the observer who experienced anger as having a better moral character than those who do not experience anger. In Studies 1-3, participants read a variety of scenarios in which an individual observed a moral violation being inflicted on a third party, and told participants that the observer either felt anger or did not feel anger before measuring participants' judgments of and behaviors toward the target. In Study 4, we tested whether presenting participants with an observer's ostensible facial expression of anger (as opposed to merely being told that the observer experienced anger) influenced evaluations of moral character, and whether expressing anger as a victim (rather than as a third-party observer) had similar reputational effects. In the final set of studies, we compared judgments about a person who experienced anger to judgments about a person who experienced sympathy (Study 5) or judgments about a person who simply believes the observed immoral action to be wrong (Studies 6a and 6b).

All sample sizes and data exclusion criteria were predetermined. We preregistered our recruitment plan, data analyses, manipulations, and all measures for Studies 2-6b, six out of the seven studies (Exp 2: https://aspredicted.org/K48_TVB; Exp 3: https://aspredicted.org/39B_J83; Exp 4: https://aspredicted.org/DVX_841; Exp 5: https://aspredicted.org/5CL_DZ1; Exp 6a: https://aspredicted.org/6CK_6KS; Exp 6b: https://aspredicted.org/3YQ_S74). In all studies, we did not recruit additional participants once we obtained our target sample sizes. We also did not

analyze data before data collection was completed. All data and materials are available at https://osf.io/nvj8x/?view_only=a86c4db4a64742b28443814b79c7a6be.

Study 1

Study 1 was designed to test the hypothesis that people would judge a target who felt angry when observing a third-party's unfair treatment more positively compared to a target who did not. We created four scenarios with different violations occurring across different contexts, including public spaces, experimental settings, and workplaces. In all scenarios, a target observed a moral violation occur and was described as having felt either angry or not in response to that violation. Participants were then asked to evaluate the target on measures of moral character and likability.

Methods

Participants

We recruited 200 U.S. participants through MTurk. We excluded participants who failed our attention check, leaving a final sample of 192 (80 women, 112 men; M_{age} = 36.89, SD_{age} = 10.59). This sample size provided 80% power to detect an effect as small as d = .04 in a two-tailed independent-samples t test with α =.05.

Design

Participants were instructed to read descriptions that were ostensibly written by other people and collected by the experimenters in a different study. We told participants that in a previous study, participants had been asked to write about a recent encounter in which they observed someone receiving unfair treatment, and that they were asked to respond to two follow-up questions regarding this experience. In fact, all the responses were created by the experimenters. Participants were randomly assigned to the *anger* condition or the *neutral*

condition. Our key manipulation—whether the person felt angry or not toward the unfair treatment they observed—was done by varying (fictitious) people's responses to one of two follow-up questions. In the *anger* condition, participants saw that the person chose "very angry" in response to the question "what was your feeling when observing what happened?". In the *neutral* condition, the person chose "not angry at all" for the same question. In both conditions participants saw that the person responded with "No, I did not say anything" for the question "did you say anything at that moment when you observe what happened?". For each condition, participants read four different scenarios, which described the experience of witnessing a third party receive unfair treatment in a dictator game, having someone cut in front of them while waiting in line, not receiving the same reward as others, and being unfairly punished (see full scenarios in the supplemental materials).

After each scenario, participants answered a series of questions regarding their impressions of the person who wrote the responses, including four semantic scales of bad-good, untrustworthy-trustworthy, unlikable-likable, and immoral-moral, and three behavioral prediction question including the extent they want to be friends with the person, the extent they want to work with the person, and the extent they think this person would act fairly toward others (all on 7-point scales). Because we were not interested in differences across the scenarios, we computed two composite scores using these responses combined across scenarios: 1) a moral character score consisting of four items within each scenario: badness/goodness, trustworthiness, morality of the target, and how fairly people believed the target would act toward others ($\alpha > .9$) and 2) a social likability score consisting of the average of three items—likability, willingness to be close friends with the target, and willingness to work on a team project together with the target ($\alpha > .9$). We used these composite scores across all studies with a focus on the moral

character score (we report the correlation between the moral character score and the social likability score in the supplemental materials).

Results and Discussion

As predicted, we found that people evaluated the angry target (M = 4.82, SD = 1.03) as having better moral character than the neutral target (M = 3.84, SD = 1.37), $t(190) = 5.60$, p =.001, d = 0.81 (see Figure 1). People also evaluated the angry target (M = 4.57, SD = 1.07) as more socially likable than the neutral target (M = 3.56, SD = 1.55), $t(190) = 5.25$, p =.001, d = 0.76.

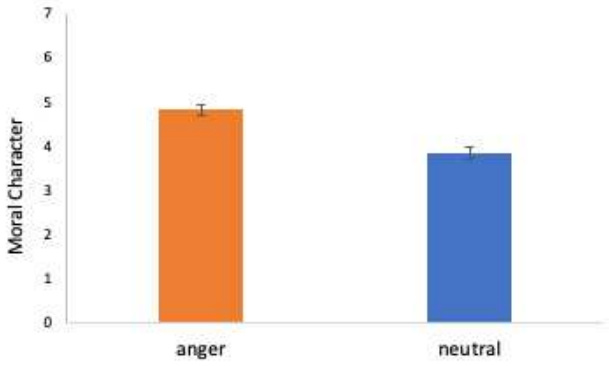

Figure 1: Moral character judgments of the target by emotional response (Study 1). Error bars are standard errors.

We also conducted an exploratory analysis that examined the mediating effect of perceived appropriateness of the response between response type and moral character ratings using a PROCESS 95% bias-corrected CI based on 5000 bootstrapped samples (Model 4, Hayes, 2012). We found that the effect of the angry target condition relative to the neutral target condition indirectly influenced the moral character judgments through higher perceived response appropriateness, ab = 0.99, 95% CI = [0.71, 1.27].

Study 1 provided support for our primary hypothesis that people would evaluate an individual who responds with anger in response to the unfair treatment of a third party more positively than an individual who does not. It is worth noting that none of the targets were described as having expressed their anger, indicating that the difference in evaluation was driven by the knowledge that the observer *felt* anger in response to the violation. In Study 2 we wanted to add to this finding by showing that the effect requires that the anger response be appropriate to the situation—feeling anger *at a moral violation* is, presumably, what gives rise to positive inferences about character, not just feeling anger.

Study 2

Study 2 was similar to Study 1, but with two main changes. First, we constructed a new set of scenarios in which an individual is described as having experienced anger in the absence of a moral violation. Second, we added a new reaction condition, where the individual was described as feeling extremely angry at what happened in the scenario. Thus, we have six different types of scenarios where there was either a moral violation or not and the target responded with feeling angry, extremely angry, or neutral. We predicted that an individual expressing unjustified anger (i.e., anger when there is no moral violation) would be judged more negatively than an individual expressing anger as a result of observing a moral violation. We also predicted that feeling *too* angry in response to a moral violation would lead to less positive moral evaluations (i.e., that a person described as "angry" would be judged more favorably than a person described as "extremely angry/furious").

Methods

Participants

We recruited 604 participants via Mturk (our target N was 600). After excluding participants who failed our attention check, we had a final sample of 600 participants (252 women, 245 men, 2 non-binary, 1 unidentified; M_{age} = 38.90, SD_{age} = 10.96). This sample size provided 80% power to detect an effect as small as f = .13 for the overall interaction with α =.05.

Design

Participants were randomly assigned to one of six between-subjects conditions, based on a 2 (behavioral type: moral violation, no violation) x 3 (response type: anger, extreme anger, neutral) design. We created new vignettes for this study that took place in different settings, including a shopping mall, a park, and a restaurant. In all conditions, participants read three vignettes describing a person observing a behavior and then that person's emotional reaction (or lack thereof) to that behavior. Each vignette described an individual observing either a *moral violation* (e.g., a shop assistant ignores waiting customers to help his friends first) or *no violation* (e.g., a shop assistant talks to his friends who are making purchase). The observer then experienced *anger* (e.g., felt angry), *excessive anger* (e.g., felt extremely angry and furious, swore at the person in his head), or felt *neutral* (e.g., didn't say anything to anyone and didn't feel anything special). For each participant, the three vignettes were consistent as to whether or not they contained a moral violation, and what kind of emotional reaction the observer experienced. After reading each scenario, participants completed the same dependent measures that were used in Study 1 with one added item directly asking participants the extent they would trust this target. This item was then included in computing the moral character score. Finally, participants completed demographic information.

Results and Discussion

To test our hypothesis that participants would have more favorable moral evaluations of an observer feeling anger only when the anger occurred as a result of a moral violation, we conducted a 2 (behavioral type: moral violation, no violation) x 3 (response type: anger, extreme anger, neutral) ANOVA on our index of moral character. The overall interaction was significant, $F(2,594) = 72.67$, $p = .001$, $\eta_p^2 = .20$ (see Figure 2). Consistent with our hypothesis, when a moral violation was present, participants judged the *angry* target ($M = 4.84$, $SD = 0.91$) as having better moral character than the *neutral* target ($M = 4.12$, $SD = 0.98$), $t(199) = -5.13$, $p = .001$, $d = 0.76$. However, when there was no moral violation, the *neutral* target ($M = 4.91$, $SD = 0.94$) was judged more positively than the *angry* target ($M = 3.68$, $SD = 1.00$), $t(199) = 8.78$, $p = .001$, $d = 1.27$.

Contrary to our initial prediction, when a moral violation was present, participants made similar evaluations of both the extremely angry target ($M = 4.81$, $SD = 1.02$) and the angry target ($M = 4.84$, $SD = 0.91$), $t(200) = 0.26$, $p = .80$, $d = 0.04$, suggesting that perhaps they did not view the extreme anger as inappropriate to the moral violation that was witnessed. However, in the absence of a moral violation, participants judged the extremely angry target ($M = 3.40$, $SD = 1.11$) as having worse moral character than the angry target ($M = 3.69$, $SD = 1.00$), $t(198) = 2.04$, $p = .04$, $d = 0.27$, suggesting that participants were not simply insensitive to the excessive anger manipulation. They found the angry person to have poor moral character, and the excessively angry person to have an even worse one.

Judgments of social likability followed a similar pattern. The overall interaction between response and behavioral type was significant, $F(2,594) = 77.40$, $p = .001$, $\eta_p^2 = .21$. In the presence of a moral violation, participants liked the angry target ($M = 4.66$, $SD = 1.00$) more than the neutral target ($M = 3.87$, $SD = 1.08$), $t(199) = 5.10$, $p = .001$, $d = 0.76$. However, in the

absence of a moral violation, people liked the neutral target ($M = 4.64$, $SD = 0.98$) more than the angry target ($M = 3.22$, $SD = 1.17$), $t(199) = 9.19$, $p = .001$, $d = 1.32$. Moreover, in the presence of a moral violation, there was no differences in evaluations of the angry target ($M = 4.66$, $SD = 1.00$) and the extremely angry target ($M = 4.56$, $SD = 1.08$), $t(200) = 0.66$, $p = .51$, $d = 0.10$. However, in the absence of a moral violation, the extremely angry target ($M = 2.82$, $SD = 1.26$) was judged as less likable than the angry target ($M = 3.22$, $SD = 1.17$), $t(198) = 2.56$, $p = .01$, $d = 0.33$.

We explored a moderated mediation model where the relationship between emotional reaction on moral character judgments through perceived appropriateness of the reaction is conditional on the behavioral type (PROCESS Model 8; Hayes, 2012). We found that behavioral type significantly moderated the indirect effect of anger relative to a neutral expression on moral character ratings through perceived appropriateness, effect = 1.98, 95% CI = [1.65, 2.33]. When there was no violation, the effect of anger relative to a neutral expression on moral character was mediated by the decrease on appropriateness, effect = -1.53, 95% CI = [-1.78, -1.29]. When there was a violation, the effect of anger relative to a neutral expression on moral character was mediated by the increase on appropriateness, effect = 0.46, 95% CI = [0.28, 0.65]; Similarly, behavioral type also moderated the indirect effect of excessive anger relative to a neutral expression on moral character ratings through perceived appropriateness, effect = 2.03, 95% CI = [1.71, 2.38]. When there was no violation, the effect of anger relative to a neutral expression on moral character was mediated by the decrease in perceived appropriateness, effect = -1.59, 95% CI = [-1.85, -1.36]. When there was a violation, the effect of anger relative to a neutral expression on moral character was mediated by the increase in perceived appropriateness, effect = 0.44, 95% CI = [0.25, 0.64].

These results suggest that people are sensitive to the appropriateness of anger such that they only evaluated the angry target more positive than the neutral target if the anger was directed at a moral violation. To our surprise, when the anger over the moral violation was described as more extreme, it did not cause more negative judgments of moral character. It could be that people simply endorse high levels of outrage over moral violations, an explanation consistent with the frequency and degree with which people express moral outrage in public life, especially when online (Crockett, 2017).

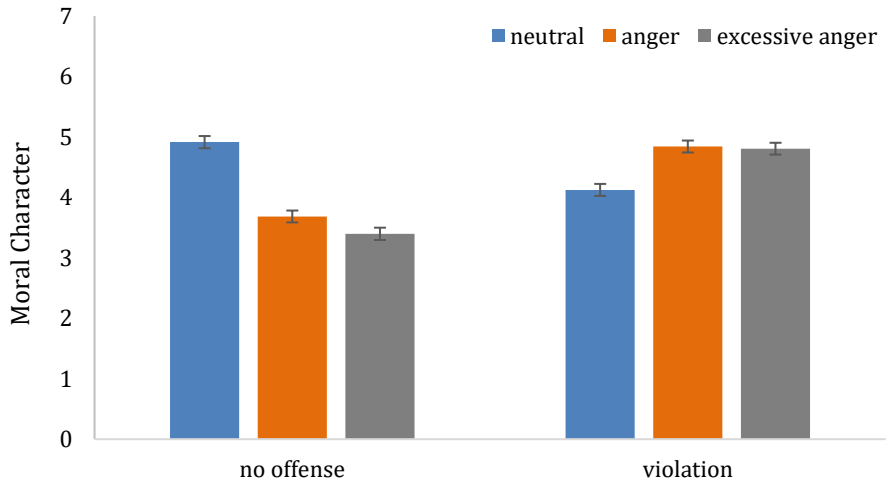

Figure 2: Moral character judgments of the target by emotional response and behavioral type (Study 2). Error bars are standard errors.

Study 3

In the previous two studies, we assessed judgments of moral character by having participants make evaluations about a target on a set of scale items. In this study we sought to test whether participants would show a difference in their behavior toward targets whom they knew felt anger over a moral violation (compared to those who did not). Building on previous

work showing that moral character information can influence behavior in a trust game (e.g., Anderson et al., 2021; Everett et al., 2016), we predicted that participants would be more willing to trust a target in a trust game (Berg et al., 1995) if that target felt anger in response to a moral violation.

Methods

Participants

We recruited 160 U.S. participants through MTurk. We based our sample size on those used in past research using a similar methodology (Anderson et al., 2021; Everett et al., 2016). Per our preregistration, we excluded participants who failed any of our three comprehension checks regarding the trust game (N = 27), leaving a final sample of 133 (61 women, 70 men, 1 non-binary, 1 left blank; M_{age} = 37.95). This sample size provided 80% power to detect an effect as small as dz = .25 in a two-tailed Wilcoxon signed-rank test with α =.05.

Design

Participants first answered open-ended questions asking how they would act in three hypothetical situations. The first situation was an adaptation of the *workplace team* scenario from Study 1, while the other two situations were filler tasks that were not relevant to our hypotheses[1]. The first situation read "Imagine that at your workplace, a team of five people completes a project. The supervisor finds a few small errors in the project, and singles-out one of the team members, publicly criticizing them and docking their pay. Even though it was a collective effort, the other team members do not receive any criticism or lose any pay. How would you feel if you saw this? Would you experience any anger?"

Participants were then introduced to the trust game (TG). In the typical TG, there are two players: an "investor" and a "trustee." The investor is endowed with some money and told that

any money they transfer (from zero to the full amount; this amount is the measure of "trust") to the trustee will be doubled, at which point the trustee can then decide to transfer a proportion of their total amount (from zero to the full amount they received) back to the investor (this amount is the measure of "trustworthiness"). After participants were given this description, they were asked three comprehension questions regarding the TG to ensure that they understood the game.

After successfully completing the comprehension questions, we then told participants that they had been assigned the role of the investor in the game, that they had been given $0.50 as their initial endowment, and that they would be playing in a trust game with one of two potential players; namely, other MTurkers who had already answered the hypothetical questions, and who had consented to sharing their answers with other participants (we stated that their own answers would not be shown to the other players). Participants were told that after they reported how they would behave in the trust game, we would randomly select one of the other players to be the participant's partner and would carry out the decisions for real, and that the participant's final bonus payment would be based on the outcomes of these decisions.

Participants were then presented (in counterbalanced order) with the responses to the *workplace team* scenario that had been ostensibly provided by the two other players who served as potential partners. Player 1 (*angry*) responded by saying "It just seems unfair to single that one person out. I think I would feel pretty angry about it." Player 2 (*non-angry*) said "It feels wrong to pick on just him, but I don't think I would feel angry." As an explicit measure of partner choice, we asked participants whom they would most prefer as a partner in the TG, Player 1 or Player 2. As indicators of trust, we asked participants how much of their $0.50 they would want to transfer if they were playing the game with Player 1, how much they would want to transfer if they were playing with Player 2 (from $0.00 to $0.50), and what percentage of

money they believed they would receive back if that particular player was their partner (from 0% to 100%).

Results and Discussion

Consistent with our predictions, as well as with the results of our previous studies, we found that participants were significantly more likely to select the partner who expressed anger at injustice (89%) than the partner who did not express anger at injustice (11%), $p < .001$.

Because the data were non-normally distributed, we used a series of Wilcoxon signed-ranks tests to compare the amount of money transferred and the percentage participants predicted they would receive in return. Supporting our hypotheses, we found that participants transferred more money to the *angry* partner ($M = 33.05$ cents) than the *non-angry* partner ($M = 19.66$ cents), $Z = 6.64$, $p > .001$, $r = .51$, and reported expecting to receive more money back from the *angry* partner ($M = 43.21\%$) than the *non-angry* partner ($M = 25.99\%$), $Z = 7.03$, $p < .001$, $r = .45$. We acknowledge that decisions regarding the amount of money transferred and the expected money received were likely influenced by the initial decision in the choice of partner. However, the nonetheless consistent pattern across all measures provides strong evidence that people are more trusting of others who experience third-party anger, and that this is reflected in their behavioral choices.

Study 4

In Studies 1-3, we focused on evaluations of targets whom we described as *feeling* anger when observing moral violations. It is possible, however, that anger's positive effect on moral character emerged because participants assumed that the anger was unexpressed (this could, for instance, mean that the person possessed more self-control, or that they did not cause any social discomfort by expressing anger outwardly). We wanted to know if our previous findings would

extend to instances in which a target actually *displayed* anger. The expression of anger has been repeatedly associated with negative dispositional inferences (Harker & Keltner, 2001; Knutson, 1996; Montepare & Dobish, 2003), so it is possible that expressing anger might trigger negative evaluations about a person's moral character. In addition, emotional expressions do not necessarily reflect the person's true inner emotional experiences, (Sasse, Spears, & Gordijn, 2018; Fischer & Evers, 2011) so it may be that the expression of the emotion is viewed as strategic, and somehow "cheaper" than genuinely experiencing the emotion (which we simply told people was the case in our previous studies). Nonetheless, we predicted that displays of anger would play a similar role in this study—that emotional expressions serve as an adequate proxy for what a person is actually feeling in response to a moral violation. In order to test this, instead of describing the target's feelings, we accompanied our vignettes with video clips that showed the target expressing anger on their face. We also included a condition in which the target expressed anger at their own unfair treatment, in order to see whether our effect was dependent on the anger being expressed as a result of a third-party violation. We predicted that expressing anger at one's own unfair treatment would not reflect positively on moral character in the same way that anger at a third-party violation (as it may be viewed as more self-interested and therefore less of a signal of positive moral character).

Methods

Participants

406 undergraduate students from a university in the northeast of the U.S. signed up and completed our study. We excluded participants who failed our attention check, leaving a final sample of 395 (288 women, 105 men, 2 non-binary; M_{age} = 20.05, SD_{age} = 1.32). This sample

size provided 80% power to detect an effect as small as $f = .14$ for the overall interaction with $\alpha = .05$.

Design

Participants were randomly assigned to one of four conditions, based on a 2 (victim of the violation: first-person or third-person) x 2 (emotional response: anger or neutral) between-subjects design. We told all participants that we had already conducted a study in which people read the same scenario (a version of the workplace punishment scenario from Study 1) that they were about to read, and who had been told to imagine themselves as the person in the scenario. We also told participants that we had recorded the previous participants' responses to the scenario on video. One group of participants read the same workplace punishment scenario from Study 1, in which a target observes the unfair treatment of another (*third-person* condition), another group of participants read a modified version in which the target themselves experiences the unfair treatment (*first-person* condition).

After each participant read the scenario (either first- or third-person), they watched a recording of a person whom they believed to be the previous participant reacting to the scenario in the other study. Participants in the *anger condition* watched a video of an individual displaying an angry expression whereas participants in the *neutral* condition watched a video of a person with a neutral expression (videos from Reed et al., 2014). Finally, participants completed the same set of items as in Study 2.

Results and Discussion

To test whether expressions of anger to first-person moral violations were perceived as less moral than anger over third-person violations, we conducted a 2 (emotional response: anger or neutral) x 2 (victim of the violation: first-person or third-person) between-subjects ANOVA

(see Figure 3). As predicted, the target who expressed anger for third-person moral violations was judged to have a more positive moral character (M = 4.85, SD = 1.01) than the neutral target (M = 3.94, SD = 0.91), $t(196) = 6.85$, p =.001, d = 0.95. When the unfair treatments were experienced first-hand, the angry target (M_{anger}= 4.65, SD_{anger}= 0.90) was also judged to have a better moral character than the neutral target ($M_{neutral}$= 4.33, $SD_{neutral}$= 0.92), $t(195) = 2.37$, p =.02, d = 0.35. Importantly, there was a significant two-way interaction between emotional response and perspective, $F(1,391) = 9.89$, p =.002, η_p^2 = .03, suggesting differential moral character judgments based on the perspective, such that the benefit invoked by anger for character judgment is larger when the anger is directed towards a third-person's experience than when the violation was toward the self.

We found a similar pattern for the social likability evaluations. The interaction between emotional response and perspective was significant, $F(1,391) = 10.39$, p =.001, η_p^2 = .03. Specifically, when the emotional response was toward other people's unfair experience, participants liked the angry target (M = 4.39, SD = 1.19) more than the neutral target (M = 3.55, SD = 1.09), $t(196) = 5.32$, p =.001, d = 0.73. However, the angry target (M = 4.02, SD = 1.00) was not liked more than the neutral target (M = 3.90, SD = 1.11) when the emotional response was a result of their first-hand experience of unfair treatment, $t(195) = 0.77$, p =.44, d = 0.11.

As in previous studies, we assessed whether perceived appropriateness of the response mediated the impact of response type on moral character judgments. Using PROCESS 95% bias-corrected CI based on 5000 bootstrapped samples, response type influenced moral character ratings through perceived appropriateness of the emotional reaction (Model 4, Hayes, 2012), ab = 0.39, 95% CI = [0.27, 0.52]. This result again suggests that perceiving anger as being the more

appropriate reaction in those situations indirectly lead to the increased positivity of moral character ratings for the angry target.

In this study we provided evidence that angry facial expressions provide much the same information about moral character as the descriptions of emotional experience in the previous studies. In addition, the positive anger effect (compared with the neutral condition) on moral character judgment diminished when anger was expressed for first person moral violations (i.e., when the target was themselves a victim), and further disappeared for the social likability judgment. These results were consistent with our previous findings that feeling anger toward a violation is perceived well by others compared with maintaining a neutral expression. However, we did see that the benefits associated with anger in the first-person condition were reduced compared with the third-person condition. This is consistent with our theorizing that anger expressed at people's own unfair treatment does not signal the same positive moral character as anger expressed at the unfair treatment of another, perhaps because it may be motivated by selfish reasons rather than other-oriented moral concerns (Baston et al., 2007; Batson et al., 2009; Cohen, Nisbett, Bowdle, & Schwarz, 1996). Yet, noticing that another person has been treated unfairly and experiencing an emotional response because of their experience could be a clear indication of prosocial tendencies. In the next study, we sought to compare the signaling value of anger to that of sympathy (viewed by many as a moral/prosocial emotion; Decety & Chaminade. 2003), as a way to test the magnitude of anger's reputational boost when compared to another moral emotion.

Figure 3: Moral character judgments of the target by emotional response and the victim of the violation (Study 4). Error bars are standard errors.

Study 5

One common emotion experienced when observing a moral transgression is that of sympathy for the victim of the transgression. Sympathy is considered by many to be a positive emotion, and experiencing sympathy may be a marker for virtue (Darwall, 1998; Eisenberg et al., 2014), as it is associated with prosocial behaviors (Eisenberg et al., 1989; Hume, 2003; Nichols, 2001). How would expressing anger toward a perpetrator of a third-party violation compare to expressing sympathy for the victim? In this study we used a similar design as in Studies 1-2, with the addition of a condition that described an individual who was described as experiencing sympathy for the victim of a transgression. We hypothesized that, while we would replicate the effect of anger leading to positive evaluations, participants would view the sympathetic target as having a better moral character (due to the fact that anger is viewed by many as an undesirable, negative emotion).

Methods

Participants

We recruited 300 U.S. participants through MTurk. We excluded participants who failed our attention check, leaving a final sample of 289 (112 women, 177 men; M_{age} = 35.19, SD_{age} = 9.78). This sample size provided 80% power to detect an effect as small as f = .18 for a main effect with α =.05.

Design

Participants in all conditions read scenarios about a target witnessing other people's unfair treatment. To diversify the type of unfair experiences people may encounter in daily lives, we created three new scenarios involving someone being accidently hit on the subway, smoking near a children's playground, and yelling at a cashier. At the end of the scenario, we described the agent's reaction to the observed unfair treatment other people experience. In the *anger* condition, we reported that the individual "didn't say anything to anyone but she felt angry". In the *neutral* condition, the individual "didn't say anything to anyone and just went about her/his business". In the *sympathy* condition, the individual "didn't say anything to anyone but she felt sympathy." Participants were randomly assigned to one of three response conditions. After reading the scenarios, participants responded to the same items as in Study 2.

Results and Discussion

To examine the effect of emotional responses on evaluations moral character, we conducted a one-way ANOVA with 3 between-subjects factors (emotional response: anger, sympathy, or neutral; see Figure 4). Character judgments were different, depending on the target's emotional responses, $F(2, 286) = 3.94$, p =.02, η_p^2 = .03. Specifically, we replicated the main finding again that people judged the angry target (M = 5.21, SD = 0.89) as possessing a

better moral character than the target who did not experience any emotions ($M = 4.79$, $SD = 1.08$), $t(186) = 2.80$, $p = .005$, $d = 0.42$. Contrary to our prediction that sympathy would lead to more positive evaluations of moral character than anger, people did not judge the character of the target who expressed sympathy ($M = 5.00$, $SD = 1.02$) as significantly different from that of the angry target ($M = 5.21$, $SD = 0.89$) or the neutral target ($M = 4.79$, $SD = 1.08$), $t(186)_{\text{anger-sympathy}} = 1.43$, $p = .15$, $d = 0.22$; $t(193)_{\text{neutral-sympathy}} = -1.43$, $p = .15$, $d = 0.20$. Interestingly, emotional responses only influenced moral character judgments but not social likability judgments. The level of likability toward the target did not differ by the target's emotional responses, $F(2, 286) = 1.49$, $p = .23$, $\eta_p^2 = .01$.

A mediation analysis using a PROCESS 95% bias-corrected CI based on 5000 bootstrapped samples (Model 4, Hayes, 2012) with the neutral condition as the reference group showed that the perceived appropriateness of the emotional reaction did not mediate the effect of the sympathy condition relative to the neutral condition on moral character ratings, $a_1b = 0.06$, 95% CI = [-0.16, 0.29] nor the effect of the anger condition relative to the neutral condition on moral character ratings, $a_2b = 0.17$, 95% CI = [-0.04, 0.40]. Different from previous studies, perceived appropriateness did not turn out to be a significant mediator of the effect of reactions on character judgments.

These results showed that anger elicited similar moral character judgments toward the target as sympathy. Although anger generally has a negative reputation whereas sympathy is considered a positive emotion, our findings showed that when an individual felt anger toward a perpetrator of a moral violation they were evaluated just as positively as an individual who felt sympathy for the victim. The fact that there was no difference between the positivity of

evaluations elicited by the two emotions seems to suggest that people do not necessarily value focusing one's moral attention on the victim more highly than focusing on the perpetrator.

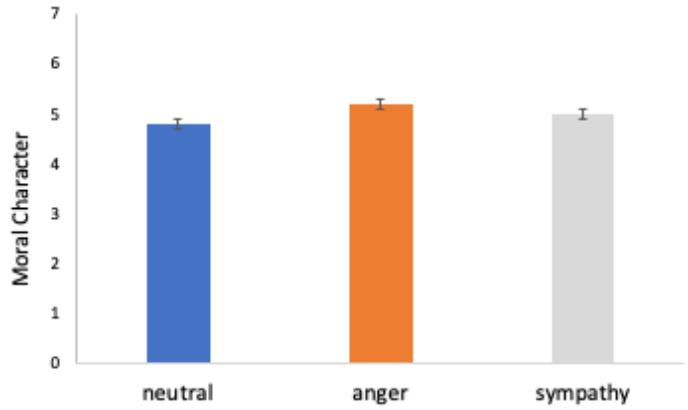

Figure 4: Moral character judgments of the target by emotional response (Study 5). Error bars are standard errors.

Study 6a

In our previous studies, we found consistent evidence that people judge an observer who experiences anger at a moral violation as having better moral character than an observer who does not experience anger. In the remaining studies, we examine the underlying mechanisms for these judgments: Why would experiencing moral anger lead to these judgments? As previously discussed, when people evaluate moral character, part of what they are evaluating is a person's "moral-cognitive machinery" (Critcher et al., 2020; Helzer & Critcher, 2018). Therefore, as a signal of moral character, morally-motivated anger should (in the minds of other people) provide information about the target's thoughts and behavioral tendencies. We examine these components in the last two studies.

One of the prerequisites for feeling angry about a moral violation is the recognition that a moral violation has occurred (Ellsworth & Smith, 1988; Montada & Schneider, 1989; Roseman,

Spindel, & Jose, 1990; van Zomeren, Postmes, & Spears, 2008). Observing someone respond with moral anger, then, may simply serve as a signal that they have recognized that an action is wrong and should be condemned. If so, this recognition alone (without the accompanying anger) should be enough to provide the reputational benefits we observed in our previous studies. However, we predict that anger provides additional information beyond the recognition that an act is wrong. For instance, anger may provide information regarding the intensity of belief that an act is wrong, or may provide information about an individual's motivation to act on their beliefs—both pieces of information that seem diagnostic of moral character. Thus, in this study, we believe that people would evaluate an angry target as having a better moral character than a target who was simply described as recognizing that an act was wrong.

Methods

Participants

We recruited 300 U.S. participants from MTurk. We excluded participants who failed our attention check, leaving a final sample of 297 (124 women, 172 men, 1 non-binary, M_{age} = 38.95, SD_{age} = 10.87). This sample size provided 80% power to detect an effect as small as f = .18 for a main effect with α =.05.

Design

In this study we used the three scenarios used in Study 2. In addition to the anger and neutral response conditions, we added a condition in which the target recognizes and acknowledges the wrongness of the transgressor's behavior (without providing information about their emotional response). Participants read all three scenarios and were randomly assigned to one of the three response conditions (anger, recognition, or neutral) before completing the same questions as in Studies 2.

Results and discussion

We conducted a one-way ANOVA with 3 (response type: anger, recognition, neutral) as a between-subject factor (see Figure 5). There was a main effect of emotional responses, $F(2, 294) = 25.18$, $p = .001$, $\eta_p^2 = .15$. Consistent with our prediction, the angry target ($M = 4.89$, $SD = 0.81$) was judged to have better moral character than the target who did not show any emotions ($M = 4.24$, $SD = 0.92$), $t(190) = 4.91$, $p = .001$, $d = 0.75$. Similarly, the target who simply recognized the act as wrong ($M = 5.14$, $SD = 1.00$) was also judged as having better moral character than the neutral target ($M = 4.24$, $SD = 0.92$), $t(198) = 6.91$, $p = .001$, $d = 0.93$. Contrary to our predictions, however, the target who simply recognized the wrongness of the act ($M = 5.14$, $SD = 1.00$) was *not* judged as having a significantly different moral character than the angry target ($M = 4.89$, $SD = 0.81$), $t(200) = 1.94$, $p = .054$, $d = 0.27$.

A one-way ANOVA with the socially likability judgments revealed a similar pattern, $F(2, 294) = 23.59$, $p = .001$, $\eta_p^2 = .14$. such that the angry target ($M = 4.72$, $SD = 0.91$) and the target who recognized the wrongness of an act ($M = 4.93$, $SD = 1.13$) were liked more than the neutral target ($M = 3.96$, $SD = 1.08$), $t(190) = 5.05$, $p = .001$, $d = 0.76$, $t(198) = 6.58$, $p = .001$, $d = 0.89$. In addition, the likability of the angry target was not significantly different from the likability for the target who recognized the act was wrong, $t(200) = 1.48$, $p = .14$, $d = 0.21$.

Using the same mediation model as in Study 5 with the neutral condition as the reference group, the effect of recognizing the act was wrong relative to the neutral condition indirectly influenced moral character ratings through increased perceived appropriateness of the reaction, $a_1b = 0.42$, 95% CI = [0.26, 0.60]. Relative to the neutral condition, the effect of the anger condition on moral character judgments was also indirectly affected by perceived

appropriateness of the reaction, $a_2b = 0.39$, 95% CI = [0.22, 0.58]. Again, replicating the mediation model with perceived appropriateness of the reaction as the mediator.

In summary, targets who were described as recognizing that an act was wrong received the same reputational benefit (compared to the neutral condition) as angry targets. These results contradicted our prediction that angry targets would be judged as having a better moral character, and seem to suggest that simply recognizing that an act is morally wrong is enough to signal a good moral character. However, it is possible that the emotion of anger nonetheless provides additional information about moral character that was not captured by our dependent variables. Given the strong association between emotions and action tendencies (Lewis, 2005; Moors et al., 2013), anger may serve as a signal of a target's motivation to act when faced with a moral violation, which itself may be morally relevant information. Finally, it is possible that these results found were simply due to a failed manipulation—participants may have simply assumed that the target who judged the action to be wrong actually felt angry, and this may have account for the similarities between moral judgments for targets in the "anger" and "recognition" conditions. In Study 6b, we sought to address these issues.

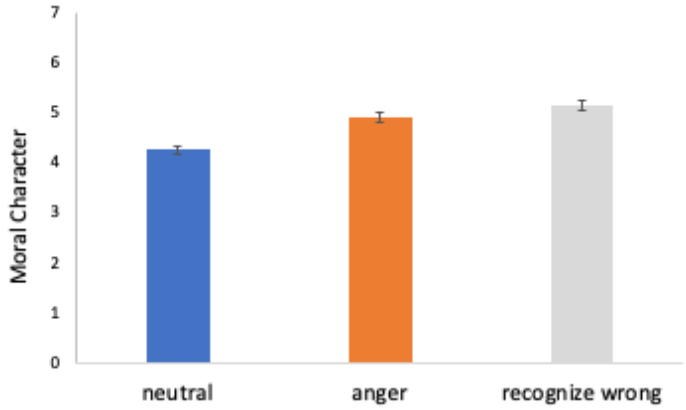

Figure 5: Moral character judgments of the target by response type (Study 6a). Error bars are standard errors.

Study 6b

In order to address the possibility that participants inferred that a target was angry when they were described as recognizing that an act was wrong, we added a condition in which we clearly stated that the target recognized the wrongness of an act but did not feel angry. In addition, we added measures intended to assess whether anger was providing relevant information above-and-beyond the recognition that an act was wrong. Given that anger has been shown to be an emotion that motivates people to act (Berkowitz, 2012; Kuppens et al., 2003; Mouilso et al., 2007), we added questions asking participants how they expected the target to act. We predicted that 1) participants would evaluate the angry target more positively than the target who recognized wrong but did not feel angry, and 2) participants would view the angry target as more motivated to act on their moral belief.

Methods

Participants

We recruited 400 U.S. participants from MTurk. We excluded two participants who failed our attention check, leaving a final sample of 398 (190 women, 198 men; M_{age} = 41.39, SD_{age} = 12.54). This sample size provided 80% power to detect an effect as small as f = .18 for a main effect with α =.05.

*D*esign

The study design was the same as Study 6a with three exceptions. First, we added a condition where we described that the target recognized that a behavior was wrong, but that he did not feel angry about it. Therefore, participants were assigned to one of four between-subjects conditions: the *angry* target, the *neutral* target, the *recognizing wrong* target, and the *recognizing wrong but not angry* target. Second, we changed our description of the target in the neutral

condition: Instead of saying the target did not feel anything special, the target was described as not thinking much about it upon seeing the event. Third, we added a question asking what participants thought about the target's likelihood to intervene in the future. For each scenario, participants were asked to imagine the target facing a similar situation in the future and indicate how likely they think the target would intervene on a 1 to 7 scale, with 1 being not likely to intervene at all and 7 being definitely going to intervene.

Results

We conducted a between-subjects one-way ANOVA (response type: anger, recognition, recognition with no anger, neutral). For moral character, we observed a main effect of response type (see Figure 6), $F(3, 394) = 10.14$, $p = .001$, $\eta_p^2 = .07$. Pairwise comparisons with Bonferroni corrections showed that, compared to the neutral target ($M = 4.03$, $SD = 1.03$), the angry target ($M = 4.71$, $SD = 0.92$; $t(198) = 4.81$, $p = .001$, $d = 0.70$), the recognizing wrong target ($M = 4.65$, $SD = 0.81$; $t(196) = 4.39$, $p = .001$, $d = 0.67$), and the recognizing wrong but not angry target ($M = 4.62$, $SD = 1.20$; $t(200) = 4.19$, $p = .001$, $d = 0.53$) were all judged as having a better moral character. Similar to our results in Study 6a, the angry target was not judged differently from the target who recognized wrong, $t(194) = 0.39$, $p = 1.00$, $d = 0.06$. However, contrary to our predictions, the angry target was *not* judged differently from the target who recognized wrong but showed no anger, $t(198) = 0.64$, $p = 1.00$, $d = 0.09$.

Judgments of likability showed the same pattern as the moral character judgments. The main effect of response type was significant, $F(3, 394) = 9.52$, $p = .001$, $\eta_p^2 = .07$. The angry target ($M = 4.48$, $SD = 0.93$; $t(198) = 4.79$, $p = .001$, $d = 0.72$), recognizing wrong target ($M = 4.35$, $SD = 0.98$; $t(196) = 3.92$, $p = .001$, $d = 0.58$), and the recognizing wrong but not angry target ($M = 4.38$, $SD = 1.27$; $t(200) = 4.16$, $p = .001$, $d = 0.53$) were all judged to be more likable

than the neutral target ($M = 3.75$, $SD = 1.10$), while the angry target was not evaluated differently from the target who recognized wrong, $t(194) = 0.85$, $p = 1.00$, $d = 0.14$, and recognized wrong and not feeling angry, $t(198) = 0.65$, $p = 1.00$, $d = 0.09$.

Based on exploratory mediational analyses as in previous studies, we found that the effect of recognizing the act was wrong without feeling anger, relative to the neutral condition, did not indirectly influence moral character ratings through perceived appropriateness of the reaction, $a_1b = 0.13$, 95% CI = [-0.06, 0.32]. However, perceived appropriateness of the reaction mediated the effect of recognizing wrong to the neutral condition and the effect of anger to the neutral condition on moral character ratings, $a_2b = 0.21$, 95% CI = [0.03, 0.40]; $a3b = 0.34$, 95% CI = [0.16, 0.53], suggesting that the reaction that was deemed as more appropriate for the situation was associated with a more positive character judgment toward the target.

To test the hypothesis that anger provided information about a target's moral motivation, we computed a composite score for the questions across all three scenarios assessing the likelihood that the target would intervene in similar situations in the future (Cronbach's alpha = 0.83) and examined whether people's expectation of the target intervening in similar future situations differed by response condition. A one-way ANOVA revealed a significant main effect of response condition, $F(3, 394) = 10.01$, $p = .001$, $\eta_p^2 = .07$ (see Figure 7). Pairwise comparisons with a Bonferroni correction showed that the angry targets ($M = 3.51$, $SD = 1.31$) were judged as most likely to intervene in similar future situations, and that this was significantly higher than the judgments of future intervention for the targets who simply recognized that a moral violation occurred ($M = 2.94$, $SD = 1.37$), $t(194) = 2.97$, $p = .02$, $d = 0.36$, for targets who recognized the moral violation and did not feel angry ($M = 2.68$, $SD = 1.34$), $t(198) = 4.40$, $p = .0001$, $d = 0.52$, and for targets in the neutral condition ($M = 2.56$, $SD = 1.31$), $t(198) = 5.03$, $p = .0001$, $d = 0.61$.

The target who was described as recognizing that an act was wrong was not evaluated differently from the neutral target $t(198) = 2.02$, $p = .26$, $d = 0.29$, or the target who recognized that an act was wrong and did not feel angry, $t(198) = 1.40$, $p = .98$, $d = 0.20$.

Discussion

In Study 6b, we replicated the findings from Study 6a such that targets who felt angry and targets who only recognized that an act was wrong were both judged to be more positive than the target in the neutral condition. This result ran contrary to our initial predictions—we thought that the target who experienced anger would be viewed as having a significantly better moral character than a target who was described as recognizing the wrongness of an act but not feeling angry at all. In other words, anger did not appear to provide an additional signal about the positivity of moral character above-and-beyond the believing that an act is wrong. This suggests that the reputational benefits of anger we have documented may be because experiencing anger communicates the moral beliefs of the target. It may be that experiencing anger in the face of a moral violation simply serves as a quick and easy-to-read signal for a third-party to figure out a target's underlying moral beliefs. This makes sense—it is likely much less common for someone observing a moral violation to explicitly state their beliefs to those around them. However, their emotional response is much more likely to be observable by surrounding witnesses.

However, consistent with our predictions, these results show that anger provides at least some information that may be morally relevant—that an angry target would be more likely to intervene in a future situation in which they observe a similar moral violation. Individuals seem to hold the belief that anger is a stronger motivator of action than simply recognizing the wrongness of an act, an explanation that is consistent with the literature on anger (Carver & Harmon-Jones, 2009). This highlights one of the beliefs that lay people hold about anger, and

this belief seem to distinguish anger from the cognitive appraisal and recognition process that precedes the emotional response.

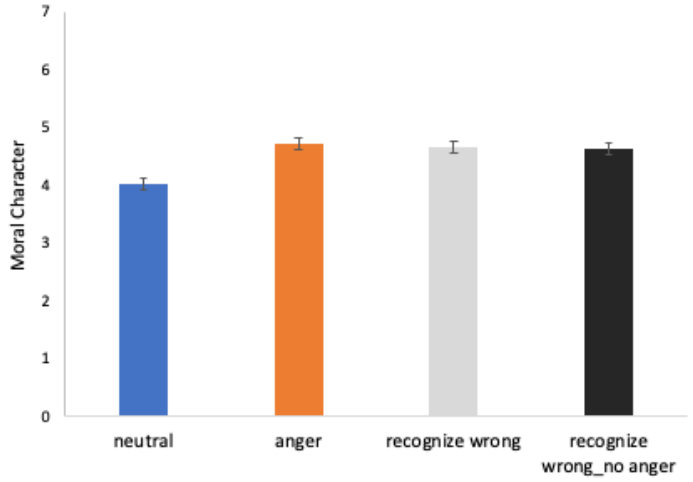

Figure 6: Moral character judgments of the target by response type (Study 6b). Error bars are standard errors.

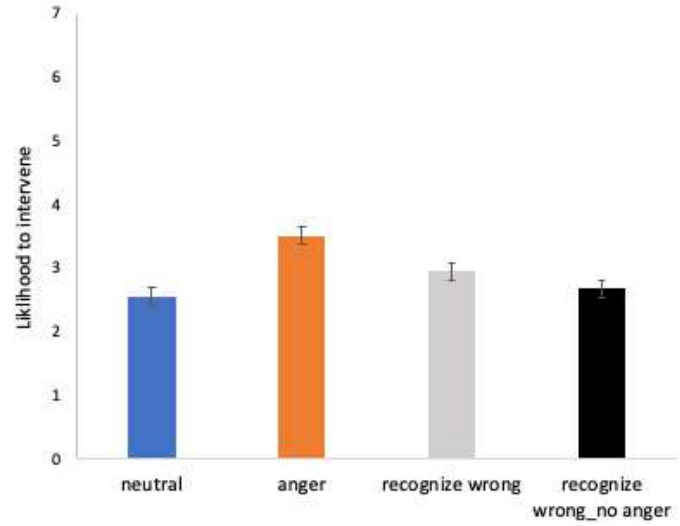

Figure 7: The target's likelihood of intervene in a future event by response type (Study 6b). Error bars are standard errors.

General Discussion

Across 7 studies employing a variety of paradigms, we found that people who experience or express anger in response to a moral transgression are judged as having a better moral

character than those who do not. In contrast to common characterizations of anger as being a purely aggressive or antisocial emotion, people actually *do* value moral anger (with some caveats). In Study 2, we demonstrated that anger is not *always* beneficial, suggesting that observers are at least partially basing their judgments on the appropriateness of the anger for the situation. We also found that these character inferences from moral anger have downstream behavioral consequences (Study 3), such that participants were more likely to choose, as a partner in a trust game, an individual who expressed anger at injustice (compared to one who did not). We also demonstrated that self-interested anger is less of a positive signal of moral character than anger when another person is harmed (Study 4). When comparing anger towards the transgressor and sympathy towards the victim as different reactions, we found that anger had an equivalent positive effect as sympathy, an altruistic positive moral emotion, on evaluations of moral character (Study 5). We also demonstrated that part of what observers may infer from anger in response to a moral violation is that the person who feels angry is likely to believe that the violation was wrong. Although we unexpectedly found that simply recognizing that an act was wrong led people to give more positive assessments of character to the same degree as the people who were described as angry, people were more likely to expect the angry target to engage in future behavioral interventions (Studies 6a and 6b).

To the best of our knowledge, this is the first set of studies that systematically looked at the sociomoral value of experiencing anger. By focusing on the emotion of anger without additional behaviors and decisions, our work serves to understand the functions of emotion, specifically anger, in the judgment of the moral character of others. Our work also extends recent theory and research on the role of moral character judgments as part of our broader social cognition (e.g., Critcher et al., 2020; Uhlmann et al., 2015). Although plenty of work has studied

how people form moral character judgments of others (e.g., Goodwin et al., 2014; Strohminger & Nichols, 2014), much of the focus has been on the role of behaviors and mental states like intentions. The present research, however, demonstrated that the experience of anger alone can be a positive signal of moral character.

Another reason that we think the positive effect of anger has on moral character is especially interesting is that anger is very often perceived as a negative emotion. However, in certain situations, which we examined in our work, showed that anger can be perceived as a positive cue for moral character judgments. One of the reasons that anger can be seen as positive is because it means that the person experiencing it can recognize immoral wrongs, and this moral recognition improves people's moral standing (Study 6a). What's more unique about anger is that people expect the angry person more likely to intervene in a future situation in which they observe a similar moral violation (Study 6b), suggesting that people hold an extra layer of expectation toward the angry targets. Different from anger, sympathy in general is considered as a positive emotion. However, the results of Study 5 showed that people did not infer more positive moral character from the person who felt sympathy, showing the strength of positivity associated with the person who experiences anger, a typical negative emotion, can be as strong as a typical positive emotion like sympathy. Although it is beyond the focus of this work, future work can further investigate when different emotions can be construed as moral and their effect on character judgments.

As mentioned above, one mechanism by which anger signals moral character appears to be by providing information that a target has judged an action to be wrong. That is, anger provides information about the person's appraisal of the situation, revealing their moral stance and beliefs. This makes sense, as appraisal processes are an integral feature of emotional

responses (Lazarus & Smith, 1988; Moors et al., 2013). Anger, specifically, is accompanied by an appraisal of injustice or unfairness (e.g., Ellsworth & Smith, 1988; Frijda et al., 1989; Mikula et al., 1998). This suggests the key element for anger to be a moral character signal is through its function of revealing the person's moral beliefs. In other words, anger may serve as a reliable and efficient cue about an individual's moral beliefs in real life (e.g., Gross et al., 1997; Hess et al., 1995; Sato & Yoshikawa, 2007). Moreover, the results from Study 6b showed that anger, compared to merely believing that an act is wrong, is more likely to communicate a motivation to act in the future. This is consistent with research on the role anger plays in motivating action, and may be why anger communicates positive character (Berkowitz, 2012; Kuppens et al., 2003; Young et al., 2011).

In addition, the exploratory mediational analyses showed that the perceived appropriateness of the response could be another mediator of the effect of response type on moral character judgments. In five out of the six mediation analyses, perceived response appropriateness significantly mediated the positive effect of anger relative to a neutral expression on moral character judgments. These results suggest that when there is a moral violation toward a third party, feeling angry toward it may be expected by people. Interestingly, such appropriateness judgments influenced how people make moral judgments of others, shedding light on the philosophical debate regarding the connection between social appropriateness and morality (Goldberg, 2022). Although this work focuses solely on the emotion of anger, we speculate a similar mediatory mechanism through perceived appropriateness to be applied to other emotions' signaling function for moral characters. For example, someone who feels disgusted by an event could be perceived as having a good moral character if the experienced disgust is deemed as appropriate for the occasion. This effect might disappear when the

experienced disgust was not considered as appropriate by the perceiver. Similarly, perceived emotional appropriateness might guide moral character judgments when the emotion is positive. For instance, someone who felt compassionate could be inferred as having a better moral character than those who don't when the experienced compassion feeling was considered as appropriate, but not when it is thought to be inappropriate. However, because these mediation analyses were exploratory, future work needs to further confirm the model and identify the key components in appropriateness (e.g., norm or expectation) that influence moral judgments.

In our studies, the boost in moral character emerged if the behavior triggering the anger was perceived to be a moral violation; if the anger was seen as unjustified, it did not lead to better character judgments. It means that, in real life, if the behavior giving rise to the outrage is not perceived as a moral violation by a large group of people, it may backfire and lead people to judge the person more negatively, or to perceive them as attempting to "virtue signal." Given that individuals differ in their perceptions of what constitutes a moral violation, often along political lines (e.g., Curry et al., 2019; Graham et al., 2009; Gray et al., 2012; Rai & Fiske, 2011), this may be the case for many such expressions of moral anger where it is expressed toward moral violations that is divided by political ideologies.

Limitations and Future Directions

Our research has several limitations that can be addressed with future research. Perhaps most critically, the majority of our studies relied on written scenarios that directly reveal the protagonist's emotional experience. This allowed us to isolate the effect we were interested in—how experiencing anger might influence observers' evaluations of moral character. However, in everyday situations, observers rarely have direct access to others' mental states in the way we described in our vignettes. Observers usually infer a target's mental states through observable

cues (Tamir & Thornton, 2018), including facial expressions (e.g., Back et al., 2009; Grossmann, 2017; Handley et al., 2021), verbal expressions (Gaudreau et al., 2015; Kraut, 1978; Shatz et al., 1983), and behavior (Kelley, 1973; Mata et al., 2021; Newtson, 1973). While we attempted testing our hypothesis with more naturalistic cues (using facial expressions in Study 4), there may still be interesting nuances in how observers infer moral character using anger expressed through different channels. Some channels might afford more extreme expression of anger than others. For example, when using language, sometimes, people might use vulgar language to express anger. However, it might put people off and compete with the positive moral inferences people make from the anger expression (i.e., anger expressed in a non-normative or overly aggressive manner might lead to less positive inferences). Future work can test the kind of anger that people deem as excessive and start reducing the positive effect associated with it.

Furthermore, in all of our studies, we used vignettes clearly depicting a specific moral violation. However, behaviors are often morally ambiguous in real life settings, or there is disagreement in the amount of harm caused by the behavior. It is unclear whether anger would confer similar reputational benefits when in response to morally ambiguous actions. Expressing anger in these cases may risk the possibility of being perceived as overly aggressive when the situation does not clearly warrant such negative emotions. In addition, if someone feels and expresses anger at *every* potential moral violation they encounter, it is possible that the reputational benefits of anger may diminish. We suspect that the effects we have documented require some amount of match between what the participant thinks about the event and what the target thinks about the event. Do both parties believe that the event was wrong and thus deserving of feeling angry about? If so, there should be reputational benefits for the target.

The generalizability of our finding to different contexts and societies is another question worth further investigation. It is possible that in real life, external factors, such as the norm or culture of the specific context, could be taken into account by people. If there is a strong norm or culture that is against the expression of anger at a specific situation, then the positive effect associated with anger could be reduced or even eliminated.

Finally, it may be that other moral emotions like disgust and contempt elicit similar effects to the ones we have documented with anger: they also signal disapproval, which could thus provide information about the person's set of beliefs. Alternatively, because disgust and contempt are associated with withdrawal and avoidance (as opposed to the approach motivation of anger), it is possible that they do not signal the same commitment to intervention in the future. So while disgust and contempt may signal the belief that an immoral act has been committed, anger may be a better signal of moral character because of its connection to taking intervening action. More broadly, this suggests that the specific inferences drawn about a person's character and traits from their emotions depends on the predicted action-patterns associated with those emotions, and not simply from whether the emotions generally reflect approval or disapproval. We believe that future studies could address these possibilities.

Conclusion

Anger has a bad reputation, and is often associated with destructive and negative outcomes. However, we provided evidence that anger experienced in the presence of a moral violation can bestow reputational rewards on those who experience it. We believe that these findings shed light on the function of anger and its social significance, as well as on our understanding of how individuals make inferences about the moral character of others.

Footnotes

[1] One filler task asked “Imagine you are walking around your town and on the sidewalk is an unmarked envelope with $100 in it. What would you do with the money?” The other filler task asked “Imagine your first cousin came to you and asked you to help cover their mortgage payment for a month. What would you do?” These filler tasks were included to increase the overall believability of the paradigm.

References

Anderson, R. A., Kamtekar, R., Nichols, S., & Pizarro, D. A. (2021). “False positive” emotions, responsibility, and moral character. *Cognition*, *214*, 104770.

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

Averill, J. R. (2012). *Anger and aggression: An essay on emotion*. Springer Science & Business Media.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American psychologist*, *38*(11), 1145-1160..

Allcorn, S. 1994. Anger in the workplace: Understanding the causes of aggression and violence. Westport, CT: Quorum Books.

Archer, J. (2000). Sex differences in aggression between heterosexual partners: a meta-analytic review. *Psychological bulletin*, *126*(5), 651-680.

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and social psychology review*, *9*(3), 212-230.

Back, E., Jordan, T. R., & Thomas, S. M. (2009). The recognition of mental states from dynamic and static facial expressions. *Visual cognition*, *17*(8), 1271-1286.

Batson, C. D., Kennedy, C. L., Nord, L. A., Stocks, E. L., Fleming, D. Y. A., Marzette, C. M., ... & Zerger, T. (2007). Anger at unfairness: Is it moral outrage?. *European Journal of Social Psychology*, 37(6), 1272-1285.

Battigalli, P., Dufwenberg, M., & Smith, A. (2019). Frustration, aggression, and anger in leader-follower games. *Games and Economic Behavior*, *117*, 15-39.

Batson, C. D., Chao, M. C., & Givens, J. M. (2009). Pursuing moral outrage: Anger at torture. *Journal of Experimental Social Psychology*, *45*(1), 155-160.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist*, *45*(4), 494-503.

Berkowitz, L. (1993). *Aggression: Its causes, consequences, and control*. Mcgraw-Hill Book Company.

Berkowitz, L. (2012). A different view of anger: The cognitive-neoassociation conception of the relation of anger to aggression. Aggressive behavior, 38(4), 322-333.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122-142.

Berger, J., & Milkman, K. L. (2012). What makes online content viral?. *Journal of marketing research*, *49*(2), 192-205.

Berman, J. Z., Levine, E. E., Barasch, A., & Small, D. A. (2015). The Braggart's dilemma: On the social rewards and penalties of advertising prosocial behavior. *Journal of Marketing Research*, *52*(1), 90-104.

Bohnert, A. M., Crnic, K. A., & Lim, K. G. (2003). Emotional competence and aggressive behavior in school-age children. *Journal of abnormal child psychology*, *31*, 79-91.

Birkley, E. L., & Eckhardt, C. I. (2015). Anger, hostility, internalizing negative emotions, and intimate partner violence perpetration: A meta-analytic review. *Clinical psychology review*, *37*, 40-56.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313-7318.

Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and social psychology bulletin*, *28*(6), 724-731.

Chuai, Y., & Zhao, J. (2022). Anger can make fake news viral online. *Frontiers in Physics*, *10*, 970174.

Crane, C. A., & Testa, M. (2014). Daily associations among anger experience and intimate partner aggression within aggressive and nonaggressive community couples. *Emotion*, *14*(5), 985.

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: evidence and implications. *Psychological bulletin*, *135*(2), 183.

Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An" experimental ethnography.". *Journal of personality and social psychology*, *70*(5), 945-960.

Critcher, C. R., Helzer, E. G., & Tannenbaum, D. (2020). Moral character evaluation: Testing another's moral-cognitive machinery. *Journal of Experimental Social Psychology*, *87*, 103906.

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4*(3), 308-315.

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature human behaviour*, *1*(11), 769-771.

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, *60*(1), 47-69.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*(4), 324-336.

Darwall, S. (1998). Empathy, sympathy, care. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 89(2/3), 261-282.

Deffenbacher, J. L., Deffenbacher, D. M., Lynch, R. S., & Richards, T. L. (2003). Anger, aggression, and risky behavior: a comparison of high and low anger drivers. *Behaviour research and therapy*, *41*(6), 701-718.

Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and cognition*, *12*(4), 577-596.

Eisenberg, N., Fabes, R. A., Miller, P. A., Fultz, J., Shell, R., Mathy, R. M., & Reno, R. R. (1989). Relation of sympathy and personal distress to prosocial behavior: a multimethod study. *Journal of personality and social psychology*, *57*(1), 55-66.

Eisenberg, N., Spinrad, T. L., & Taylor, Z. E. (2014). Sympathy. In *The Handbook of Virtue Ethics* (pp. 417-425). Routledge.

Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. Motivation and Emotion, 12, 271-302.

Eisenberg, N., Spinrad, T. L., & Taylor, Z. E. (2014). Sympathy. In The Handbook of Virtue Ethics (pp. 417-425). Routledge.

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*(6), 772-787.

Fischer, A. H., Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. Journal of Personality and Social Psychology, 93, 103–115.

Fischer, A. H., & Evers, C. (2011). The social costs and benefits of anger as a function of gender and relationship context. *Sex roles*, *65*, 23-34.

Fernandez, E. & Kerns, R. D. (2008). Anxiety, depression, and anger: The core of negative affect in medical populations. InG.J.Boyle,D.Matthews&D.Saklofske(Eds.), International Handbook of Personality Theory and Testing: Vol. 1: Personality Theories and Models (pp. 659–676). London: Sage.

Fedotova, N. O., Fincher, K. M., Goodwin, G. P., & Rozin, P. (2011). How much do thoughts count?: Preference for emotion versus principle in judgments of antisocial and prosocial behavior. *Emotion Review*, *3*(3), 316-317.

Frijda, N. H., Kuipers, P., & Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of personality and social psychology*, *57*(2), 212.

Friedman, R., Anderson, C., Brett, J., Olekalns, M., Goates, N., & Lisco, C. C. (2004). The positive and negative effects of anger on dispute resolution: evidence from electronically mediated disputes. *Journal of Applied Psychology*, *89*(2), 369.

Gaudreau, G., Monetta, L., Macoir, J., Poulin, S., Laforce Jr, R., & Hudon, C. (2015). Mental state inferences abilities contribution to verbal irony comprehension in older adults with mild cognitive impairment. *Behavioural neurology*, *2015*.

Gibson, D. E., & Callister, R. R. (2010). Anger in organizations: Review and integration. *Journal of management*, *36*(1), 66-93.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of personality and social psychology*, *106*(1), 148.

Gold, G. J., & Weiner, B. (2000). Remorse, confession, group identity, and expectancies about repeating a transgression. *Basic and applied social psychology*, *22*(4), 291-300.

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, *23*(2), 206-215.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, *96*(5), 1029-1046.

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam Skorpen, C., & Hsu, A. Y. (1997). Emotion and aging: experience, expression, and control. *Psychology and aging*, *12*(4), 590-599.

Grossmann, T. (2017). The eyes as windows into other minds: An integrative perspective. *Perspectives on Psychological Science*, *12*(1), 107-121.

Hartman, R., Blakey, W., & Gray, K. (2022). Deconstructing moral character judgments. *Current Opinion in Psychology*, *43*, 205-212.

Hess, U., Banse, R., & Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of personality and social psychology*, *69*(2), 280-288.

Handley, G., Kubota, J. T., Li, T., & Cloutier, J. (2021). Impact of interracial contact on inferring mental states from facial expressions. *Royal Society Open Science*, *8*(7), 202137.

Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Morality's centrality to liking, respecting, and understanding others. *Social Psychological and Personality Science*, *7*(7), 648-657.

Helen Landmann & Ursula Hess (2017) What elicits third-party anger? Theeffects of moral violation and others' outcome on anger and compassion, Cognition and Emotion,31:6, 1097-1111.

Heiphetz, L., Strohminger, N., Gelman, S. A., & Young, L. L. (2018). Who am I? The role of moral beliefs in children's and adults' understanding of identity. *Journal of Experimental Social Psychology*, *78*, 210-219.

Helzer, E. G., & Critcher, C. R. (2018). What do we evaluate when we evaluate moral character. *Atlas of moral psychology*, 99-107.

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In *Advances in experimental social psychology* (Vol. 30, pp. 1-46). Academic Press.

Hoffman, M. L. (1990). Empathy and justice motivation. *Motivation and emotion*, *14*, 151-172.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York: Cambridge University Press.

Hume, D. (2003). A treatise of human nature. Courier Corporation.

Izard, C. E. (1991). *The psychology of emotions*. Springer Science & Business Media.

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, *530*(7591), 473-476.

Jordan, J. J., & Rand, D. G. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of personality and social psychology*, *118*(1), 57-88.

Kassinove, H., & Tafrate, R. C. (2002). *Anger management: The complete treatment guidebook for practitioners*. Impact Publishers.

Karasawa, K. (2001). Anger vs guilt: Inference of responsibility attribution and interpersonal reactions. *Psychological Reports*, *89*(3), 731-739.

Keltner, D., & Lerner, J. S. (2010). Emotion.

Kelley, H. H. (1973). The processes of causal attribution. *American psychologist*, *28*(2), 107-128.

Kish-Gephart, J. J., Detert, J. R., Treviño, L. K., & Edmondson, A. C. (2009). Silenced by fear:: The nature, sources, and consequences of fear at work. *Research in organizational behavior*, *29*, 163-193.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *journal of Nonverbal Behavior*, *20*(3), 165-182.

Kraut, R. E. (1978). Verbal and nonverbal cues in the perception of lying. *Journal of personality and social psychology*, *36*(4), 380-391.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human behavior*, *28*(2), 75-84.

Kuppens, P., Van Mechelen, I., Smits, D. J. M., & De Boeck, P. (2003). The appraisal basis of anger: Specificity, necessity and sufficiency of components. *Emotion, 3*, 254–269.

Lazarus, R. S., & Smith, C. A. (1988). Knowledge and appraisal in the cognition—emotion relationship. *Cognition & Emotion, 2*(4), 281-300.

Larsson, J., Bjureberg, J., & Hesser, H. (2023). Anger Profiles Among Individuals Seeking Treatment for Maladaptive Anger: Associations with Emotion Regulation. *Clinical Psychology & Psychotherapy*. 2023, Vol.30(3), p.599-610

Landmann, H., & Hess, U. (2017). What elicits third-party anger? The effects of moral violation and others' outcome on anger and compassion. *Cognition and emotion*, *31*(6), 1097-1111.

Lee, A. H., & DiGiuseppe, R. (2018). Anger and aggression treatments: a review of meta-analyses. *Current opinion in psychology*, *19*, 65-74.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of personality and social psychology*, *93*(2), 234-249.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & emotion*, *14*(4), 473-493.

Lerner, J. S., & Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of behavioral decision making*, *19*(2), 115-137.

Lotz, S., Okimoto, T. G., Schlösser, T., & Fetchenhauer, D. (2011). Punitive versus compensatory reactions to injustice: Emotional antecedents to third-party interventions. *Journal of experimental social psychology*, *47*(2), 477-480.

Iyer, A., Schmader, T., & Lickel, B. (2007). Why individuals protest the perceived transgressions of their country: The role of anger, shame, and guilt.Personality and Social Psychology Bulletin, 33, 572–587

Mata, A., Vaz, A., & Palma, T. A. (2021). Cue competition in mental state inference: Blocking effects in learning to interpret the behaviors of others. *Journal of Experimental Social Psychology*, *95*, 104122.

Martin, R., Wan, C. K., David, J. P., Wegner, E. L., Olson, B. D., & Watson, D. (1999). Style of anger expression: Relation to expressivity, personality, and health. *Personality and Social Psychology Bulletin*, *25*(10), 1196-1207.

Matsumoto, D., Yoo, S. H., & Chung, J. (2010). The expression of anger across cultures. *International handbook of anger: Constituent and concomitant biological, psychological, and social processes*, 125-137.

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and social psychology bulletin*, *24*(7), 769-783.

Montada, L., & Schneider, A. (1989). Justice and emotional reactions to the disadvantaged. *Social Justice Research*, 3, 313–344.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal behavior*, *27*, 237-254.

Mouilso, E., Glenberg, A. M., Havas, D. A., & Lindeman, L. M. (2007, August). Differences in action tendencies distinguish anger and sadness after comprehension of emotional sentences. In Proceedings of the 29th annual cognitive science society (Vol. 14, pp. 1325-1330). Austin, TX.. Cognitive Science Society.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119-124.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of personality and social psychology*, *28*(1), 28-38.

Nichols, S. (2001). Mindreading and the cognitive architecture underlying altruistic motivation. Mind & language, 16(4), 425-455.

Norlander, B., & Eckhardt, C. (2005). Anger, hostility, and male perpetrators of intimate partner violence: A meta-analytic review. *Clinical psychology review*, *25*(2), 119-152.

Reed, L. I., DeScioli, P., & Pinker, S. A. (2014). The commitment function of angry facial expressions. *Psychological Science*, *25*(8), 1511-1517.

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological science*, *14*(3), 267-272.

Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological review*, *118*(1), 57-75.

Rafaeli, A., & Sutton, R. I. (1989). The expression of emotion in organizational life. *Research in organizational behavior*, *11*(1), 1-42.

Reed, L. I., DeScioli, P., & Pinker, S. A. (2014). The commitment function of angry facial expressions. *Psychological Science*, *25*(8), 1511-1517.

Roberton, T., Daffern, M., & Bucks, R. S. (2012). Emotion regulation and aggression. *Aggression and violent behavior*, *17*(1), 72-82.

Roberton, T., Daffern, M., & Bucks, R. S. (2015). Beyond anger control: Difficulty attending to emotions also predicts aggression in offenders.*Psychology of Violence, 5*(1), 74–83.

Robinson, M. D., Boyd, R. L., & Persich, M. R. (2016). Dispositional anger and the resolution of the approach–avoidance conflict. *Emotion*, *16*(6), 838-849.

Roseman, I. J., Spindel, M. S., & Jose, P. E. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of Personality and Social Psychology, 59*, 899-915.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological bulletin*, *110*(3), 426-450.

Sato, W., & Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, *104*(1), 1-18.

Sasse, J., Spears, R., & Gordijn, E. H. (2018). When to reveal what you feel: How emotions towards antagonistic out-group and third party audiences are expressed strategically. *Plos one*, *13*(9), e0202163.

Scarpa, A., & Raine, A. (2000). Violence associated with anger and impulsivity. *The neuropsychology of emotion*, 320-339.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, *14*(3), 301-321.

Sinaceur, M., & Tiedens, L. Z. (2006). Get mad and get more than even: When and why anger expression is effective in negotiations. *Journal of Experimental Social Psychology*, *42*(3), 314-322.

Sinaceur, M., Van Kleef, G. A., Neale, M. A., Adam, H., & Haag, C. (2011). Hot or cold: Is communicating anger or threats more effective in negotiation?. *Journal of Applied Psychology*, *96*(5), 1018.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159-171.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in cognitive sciences*, *22*(3), 201-212.

Toribio-Flórez, D., Saße, J., & Baumert, A. (2023). "Proof under reasonable doubt": Ambiguity of the Norm Violation as Boundary Condition of Third-Party Punishment. *Personality and Social Psychology Bulletin*, *49*(3), 429-446.

Tracy, J. L., & Robins, R. W. (2007). Emerging insights into the nature and function of pride. *Current directions in psychological science*, *16*(3), 147-150.

Tiedens, L. Z. (2001). The effect of anger on the hostile inferences of aggressive and nonaggressive people: Specific emotions, cognitive processing, and chronic accessibility. Motivation and Emotion, 25(3), 233–251

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72-81.

Van Doorn, J., Zeelenberg, M., & Breugelmans, S. M. (2014). Anger and prosocial behavior. *Emotion Review*, *6*(3), 261-268.

Van Doorn, J., Zeelenberg, M., Breugelmans, S. M., Berger, S., & Okimoto, T. G. (2018). Prosocial consequences of third-party anger. *Theory and Decision*, *84*, 585-599.

Van de Vyver, J., & Abrams, D. (2015). Testing the prosocial effectiveness of the prototypical moral emotions: Elevation increases benevolent behaviors and outrage increases justice behaviors. *Journal of Experimental Social Psychology*, *58*, 23-33.

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. Psychological Bulletin, 134, 504-535.

van Honk, J., Schutter, D. J. L. G. (2007). Vigilant and avoidant responses to angry facial expressions: Dominance and submission motives. In Harmon-Jones, E., Winkielman, P. (Eds.), Social neuroscience: Integrating biological and psychological explanations of social behavior (pp. 197–223). New York, NY: Guilford Press.

Van Kleef, G. A., De Dreu, C. K., & Manstead, A. S. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of personality and social psychology*, *86*(1), 57-76.

Van Kleef, G. A., Homan, A. C., Beersma, B., Van Knippenberg, D., Van Knippenberg, B., & Damen, F. (2009). Searing sentiment or cold calculation? The effects of leader emotional displays on team performance depend on follower epistemic motivation. *Academy of Management Journal*, *52*(3), 562-580.

Van Kleef, G. A., & De Dreu, C. K. (2010). Longer-term consequences of anger expression in negotiation: Retaliation or spillover?. *Journal of Experimental Social Psychology*, *46*(5), 753-760.

Veenstra, L., Schneider, I. K., Bushman, B. J., & Koole, S. L. (2017). Drawn to danger: Trait anger predicts automatic approach behaviour to angry faces. *Cognition and Emotion*, *31*(4), 765-771.

Vitaglione, G. D., & Barnett, M. A. (2003). Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. *Motivation and Emotion*, *27*, 301-325.

Waldman, D. A., Balthazard, P. A., & Peterson, S. J. (2011). Leadership and neuroscience: Can we revolutionize the way that inspirational leaders are identified and developed?. *Academy of Management Perspectives*, *25*(1), 60-74.

Wakslak, C. J., Jost, J. T., Tyler, T. R., & Chen, E. S. (2007). Moral outrage mediates the dampening effect of system justification on support for redistributive social policies. *Psychological science*, *18*(3), 267-274.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, *92*(4), 548.

Westra, E. (2021). Virtue signaling and moral progress. *Philosophy & Public Affairs*, *49*(2), 156-178.

Young, M. J., Tiedens, L. Z., Jung, H., & Tsai, M. H. (2011). Mad enough to see the other side: Anger and the search for disconfirming information. *Cognition and Emotion*, *25*(1), 10-21.