

* THE CURRENT WORK HAS NOT BEEN PEER-REVIEWED*

Being Good to Look Good:
Moral Character Is Positively Associated with Hypocrisy Among Reputation-Seeking
Individuals

Mengchen Dong^{1,2*}, Tom R. Kupfer¹, Jan-Willem Van Prooijen¹

¹ Vrije Universiteit Amsterdam

² Max Planck Institute for Human Development

* To whom correspondence should be addressed: dong@mpib-berlin.mpg.de

Abstract

Moral character is widely believed to guide a moral and prosocial life, navigating individuals through decisions about right or wrong. People with a strong moral character therefore may not be expected to behave hypocritically, by imposing stringent moral standards on others but not on themselves. But from an evolutionary perspective, moral character partly functions to maintain a positive reputation, prompting a motivation to *appear* moral. This account does predict a positive association between moral character and hypocrisy, particularly for individuals who are strongly motivated to gain a positive reputation. Three studies (employing vignettes, large-scale multinational panel data, and a behavioral experiment) revealed that various indicators of moral character (justice sensitivity, moral value, and moral identity) predicted harsher judgments of others' than own transgressions. These self-other discrepancies emerged particularly when people possessed strong reputation management motives. The findings highlight how reputational concerns moderate the link between moral character and moral judgment.

Keywords: morality; moral character; hypocrisy; reputation; status

Being Good to Look Good:

Moral Character Is Positively Associated with Hypocrisy Among Reputation-Seeking Individuals

From Aristotle's virtue ethics onwards, philosophers, social scientists, and educators have deemed that moral character paves the way to moral judgments and practices (Aquino & Reed, 2002; Kamtekar, 2004; Walker et al., 1987; Walker & Frimer, 2007). The development of moral character, including, for example, sensitivity to justice-related incidents, endorsement of moral values, and identification with morality as part of self-concept, is seen as a fundamental goal of civic education (Althof & Berkowitz, 2006; Walker et al., 1987). Individuals with a strong moral character therefore may be expected to impose the same moral standards on themselves as on others. In other words, they should not be moral hypocrites.

From an evolutionary perspective, however, one key function of being a moral person is to maintain a moral reputation, thereby attracting cooperators and avoiding social exclusion (Nowak & Sigmund, 2005; Sperber & Baumard, 2012; Vonasch et al., 2018). One of the best ways to acquire and maintain a moral reputation is to internalize a strong moral character (Heintz et al., 2016; Trivers, 1971). Facing various moral tests in daily life, moral character can prompt people to act in socially desirable ways (Jordan et al., 2016; Jordan & Rand, 2020; Paulhus, 1984), and serve to appear moral to others (Batson et al., 1997, 1999; Lönnqvist et al., 2015; Shaw et al., 2014). We propose that people with a strong moral character can demonstrate moral hypocrisy—which we conceptualize as stronger disapproval

of others' rather than their own misdeeds (Graham et al., 2015; Lammers, 2012; Polman & Ruttan, 2012; Valdesolo & DeSteno, 2008; Weiss et al., 2018)—especially when they are strongly motivated to maintain a good reputation. Below we present our reasoning in greater detail.

People who have a strong moral character, by definition, are motivated by prosocial motives such as compassion (Reed & Aquino, 2003) and guilt proneness (Cohen et al., 2012). But from an evolutionary perspective, having a strong moral character partially serves a reputation management function (Sperber & Baumard, 2012; Trivers, 1971). To illustrate, a person's moral identity, as measured with self-reported importance of several moral traits (caring, compassionate, fair, friendly, generous, helpful, hardworking, honest, and kind), does not only reflect prosocial motives but also “the importance of maintaining and conveying a socially constructed image of a moral person” (Reed & Aquino, 2003, p. 1272). We suggest that if moral character partially serves a reputation management function, people with a strong moral character may condemn others' misdeeds more harshly than their own. This is because people can appear moral to others either by enhancing their own good deeds (Hardy & Van Vugt, 2006; Jordan et al., 2016) or by condemning others' misdeeds (Jordan et al., 2017; Simpson et al., 2013). Meanwhile, outrage over others' transgressions can be a relatively costless way to signal one's virtues and enhance one's own moral reputation (Crockett, 2017; Jordan et al., 2017; Jordan & Rand, 2020; Simpson et al., 2013).

Previous studies show that people with a strong moral character can behave in strategic ways that serve a good reputation. For example, older (vs. younger) children—who

usually have received more years of character/civic education, and may be expected to be ahead in their moral development—are not only more likely to distribute resources fairly, but are also more sophisticated at disguising their selfish choices with fair appearances (e.g., Shaw et al., 2014). Moreover, moral identity predicts public moral claims but not private acts of integrity (Dong et al., 2019). Likewise, a meta-analysis found that moral identity was strongly correlated with self-reported measures, but not actual observations, of moral behaviors (Aquino & Reed, 2002; Hertz & Krettenauer, 2016). Finally, environmentalist identity predicted pro-environmental behaviors, but the effects were stronger for high-visibility (e.g., carrying reusable bags) rather than low-visibility activities (e.g., saving household energy; Brick et al., 2017). The above studies suggest a positive relation between moral character and reputation-serving behaviors. Moral character therefore may predict hypocrisy, through which people can signal their moral goodness by condemning others' misdeeds without imposing equivalent standards on themselves.

We do not suggest that people place high importance in moral character *only* to appear moral, but as a product of *both* “genuine” moral motives, and the desire for a good moral reputation (Reed & Aquino, 2003; Trivers, 1971). Importantly, the relative strength of these motives is likely to vary between people. One individual might have a strong moral character while caring little for their own reputation, whereas another individual might cultivate a strong moral character primarily for the social benefits of a good reputation. The latter case would suggest an association between moral character and hypocrisy particularly among individuals motivated by reputation management. Supporting our proposition, people

higher on reputation management motives are more attentive to their self-presentation (Blader & Chen, 2012; Flynn et al., 2006; Griskevicius et al., 2010). For example, motivated by reputational concerns, people make environmentally friendly purchases more in public than in private (Griskevicius et al., 2010). We therefore predict that people with a strong (vs. weak) moral character judge others more harshly than themselves especially when they are strongly (vs. weakly) motivated by reputation management.

The Current Research

In the current research, three studies examined the relationship between moral character and moral hypocrisy, depending on the actors' reputation management motives. Moral character can manifest in various forms including, for example, sensitivity to justice-related incidents, endorsement of moral values, and identification with morality as part of self-concept. While previous studies often focus on one particular form of moral character (e.g., Schmitt et al., 2005; Hertz & Krettenauer, 2016), to test the generalizability of our proposition, we operationalized moral character differently across the three studies, respectively as justice sensitivity, moral values, and moral identity. Consistent with previous studies (e.g., Anderson et al., 2015; Flynn et al., 2006; Griskevicius et al., 2010), we operationalized reputation management motives as motives to attain social esteem and status. More specifically, reputation management motives were measured with people's self-monitoring of socially desirable behaviors, self-importance of reputation/status values, and their inherent concern for social status. Moreover, the predictions on moral hypocrisy were tested through various methods, including vignettes (Study 1), large-scale multinational panel

data (Study 2), and a behavioral experiment (Study 3). As such, the present study sought to establish converging evidence for our propositions across multiple conceptualizations of moral character and reputation management motives, while using complementary research methods.

From an individual difference perspective, we expect a three-way interaction effect among moral character, reputation management motives, and moral target, such that moral character predicts stronger moral condemnation of transgressions committed by others than oneself (i.e., moral hypocrisy), which is especially true for people with strong (vs. weak) reputation management motives. Across the three studies, we reported all the materials, measures, and exclusions (if any). Study 2 used existing archived data; except for that, Studies 1 and 3 determined the sample size before data collection and any analyses.

Study 1

Study 1 tested moral hypocrisy in the blame of own versus others' hypothetical transgressions, as a function of moral character and reputation management motives. We operationalized moral character as dispositional justice sensitivity (Schmitt et al., 2005), which is often regarded as a central component of moral character (Kohlberg, 1976), and predicts both own moral deeds (Fetchenhauer & Huang, 2004) and moral harshness towards others (Yoder & Decety, 2014). Reputation management motives were captured by self-monitoring, which depicts the propensity to modify behaviors depending on anticipated social approval (Flynn et al., 2006; Snyder, 1974).

Method

Participants. An average effect size of social psychological studies ($\eta_p^2 = .04$; Richard et al, 2003) required a sample size of $N = 191$, to detect our intended moral character by status motive by moral target interaction effect with 80% power at an alpha level of 0.05. Eventually, 198 participants (46 males; $M_{\text{age}} = 22.29$, $SD = 3.84$; 80.3% as students) completed our survey on SurveySwap—an online platform where people publish and complete surveys as an exchange. We retained all their data in further analyses.

Design and materials. Participants were randomly assigned to either a self-as-target or other-as-target condition. In both conditions, participants followed the same procedure: They read and evaluated four transgressive scenarios (see Supplementary Materials [SM]; adapted from Lammers, 2012; Weiss et al., 2018), and then completed the measures of justice sensitivity and self-monitoring.

In the self-as-target condition, we described four organizational scenarios and asked participants to imagine themselves enacting questionable behaviors in these scenarios (e.g., sharing information about a confidential project with a friend). After reading each scenario, participants answered four questions about moral blame on a 7-point scale ranging from 1 = *not at all* to 7 = *absolutely* (e.g., “It would reflect poorly on me if I shared information about this confidential project with my friend”; as in Weiss et al., 2018; $\alpha = 0.65$ across 16 items). In the other-as-target condition, participants read about their “co-worker” committing identical transgressions and indicated their judgments on equivalent items ($\alpha = 0.71$ across 16 items).

Participants then indicated their justice sensitivity. The original scale comprised three dimensions of perpetrator sensitivity (benefiting from unjust events), observer sensitivity (observing others being treated unjustly), and victim sensitivity (experiencing injustice towards oneself). We employed the first two dimensions given their particular relevance for people's moral character. Twenty questions (ten on perpetrator sensitivity, e.g., "It worries me for a long time when others have to fix my carelessness", $\alpha = 0.89$; ten on observer sensitivity, e.g., "I am upset when someone does not get a reward he/she has earned", $\alpha = 0.81$) were rated on a 7-point scale (from 1 = *Not at all* to 7 = *Exactly*) and were aggregated as an overall indicator of justice sensitivity ($\alpha = 0.90$).

We measured participants' reputation management motives with their total score on the self-monitoring scale (e.g., "I guess I put on a show to impress or entertain people"; $\alpha = 0.58$ across 25 binary items; Snyder, 1974).

Results

In a multivariate linear regression model, we regressed moral blame ($M = 4.54$, $SD = 0.75$) on target (self = -1, other = 1), mean-centered justice sensitivity ($M = 3.15$, $SD = 0.89$) and self-monitoring ($M = 12.08$, $SD = 3.67$), their two-way interactions, and the three-way interaction. The results revealed that justice sensitivity was positively associated with overall moral blame ($B = 0.18$, $SE = 0.06$, $t = 2.89$, $p = .004$, $\eta_p^2 = .03$, 95%CI [0.06, 0.30]), and the effect was further moderated by self-monitoring ($B = -0.04$, $SE = 0.02$, $t = -2.04$, $p = .04$, $\eta_p^2 = .03$, 95%CI [-0.07, -0.001]), such that the positive relationship between justice sensitivity

and moral blame only emerged among low ($B = 0.33$, $SE = 0.09$, $t = 3.49$, $p < .001$) but not high ($B = 0.03$, $SE = 0.09$, $t = 0.37$, $p = .71$) self-monitors.

More importantly, as predicted, the three-way interaction of justice sensitivity, self-monitoring, and moral target was significant ($B = 0.08$, $SE = 0.02$, $t = 4.18$, $p < .001$, $\eta_p^2 = .08$, 95% CI [0.04, 0.11]). As shown in Figure 1, simple slope analyses revealed that justice sensitivity predicted stronger blame of own transgressions, but only for participants low (-1 SD; $B = 0.60$, $SE = 0.14$, $t = 4.36$, $p < .001$), not high ($+1$ SD; $B = -0.25$, $SE = 0.11$, $t = -2.21$, $p = .03$) on self-monitoring. In contrast, a reversed pattern emerged in blaming others' transgressions, such that justice sensitivity predicted harsher blame of others, when people were high ($+1$ SD; $B = 0.33$, $SE = 0.14$, $t = 2.40$, $p = .02$) rather than low (-1 SD; $B = 0.04$, $SE = 0.13$, $t = 0.33$, $p = .74$) on self-monitoring. None of the other effects were significant ($ps > .13$, $\eta_p^2 < .01$).

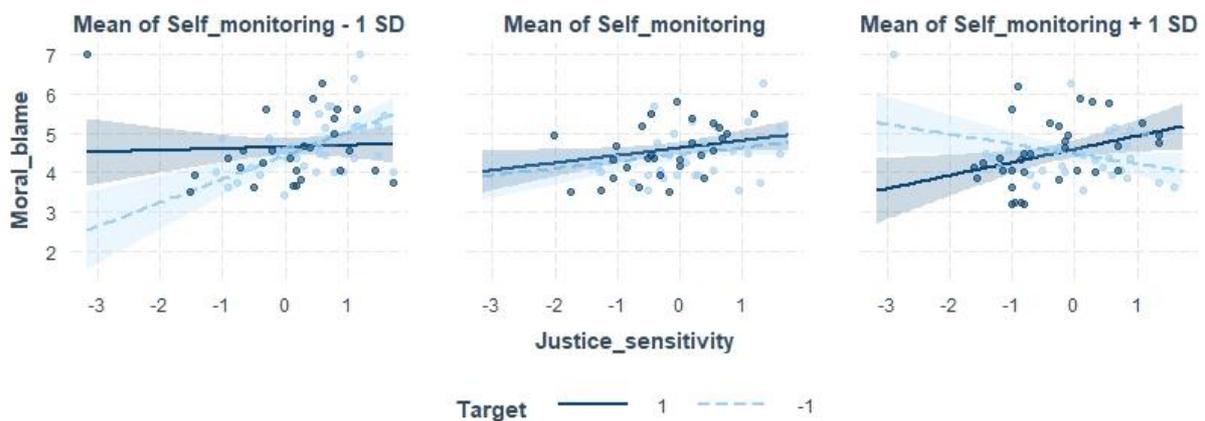


Figure 1. Moral blame of own (= -1) versus others' (= 1) hypothetical transgressions, as a function of justice sensitivity and self-monitoring in Study 1.

Discussion

Study 1 results supported the interaction effect of moral character and reputation management motives on moral hypocrisy. When they had a tendency to monitor their self-presentation in the eyes of others, people reporting more (vs. less) sensitivity to justice were more likely to blame others, but downplay their own transgressions in hypothetical scenarios.

Study 2

Findings in Study 1 are restricted given its sample composition of mainly students (80.3%) and females (76.7%). Relatedly, students' imagination of organizational scenarios and their "co-worker" may not reflect their real-life moral decisions. Therefore, Study 2 aimed to replicate Study 1 findings with a more representative sample. We employed data from the second-round European Social Survey (ESS; 2004), in which transgressive behaviors were indicated by real-life wrongdoings (e.g., making an exaggerated insurance claim). Moreover, we conceptualized people's moral character and reputation management motives as their personal identifications with relevant values (i.e., moral value and reputation/status value respectively).

Method

Participants. We used data from the 2004-wave of the ESS, which had a unique module investigating people's perception and enactment of economic morality. The 2004-wave collected data of 47,537 participants from 25 countries via face-to-face interviews. The

current research included 36,369 participants (17,006 males, 19,310 females, and 53 missing; $M_{\text{age}} = 46.24$, $SD = 17.87$), who gave valid answers to all the examined questions¹.

Measures. For each respondent, the ESS 2004 included items examining both frequencies of own economic transgressions and judgments of others' economic transgressions. Participants first indicated the moral wrongness of others' transgressions, including "someone paying cash without receipt to avoid VAT or tax", "someone selling something second-hand and concealing faults", "someone making exaggerated/false insurance claim" and "public official asking favor/bribe in return for service" (on a 4-point scale from 1 = *Not wrong at all* to 4 = *Seriously wrong*; $\alpha = 0.69$). After responding to some unrelated questions, participants then indicated their own transgression frequencies in the last five years², including "sold something second-hand and concealed its faults" "misused/altered card/document to pretend eligible" "made an exaggeration or false insurance claim" "offered favor/bribe to public official for service" and "falsely claim government benefit: social security or other" (on a scale with 1 = *Never*, 2 = *Once*, 3 = *Twice*, 4 = 3 to 4 times, and 5 = *5 times or more*; $\alpha = 0.54$).

The survey captured moral value ("important to help people and care for others' well-being" and "important to behave properly"; $r = .27$, $p < .001$) and reputation/status value ("important to show abilities and be admired" and "important to get respect from others"; $r = .38$, $p < .001$) by having participants indicate the extent to which a hypothetical person with

¹ We excluded respondents whose answers were recorded as "Refusal" "Don't know" or "No answer" on any of the targeted questions.

² Another item "Kept change from shop assistant/waiter when given too much" was not included given the low reliability coefficient of the six-item measure ($\alpha = 0.27$).

the aforementioned values was like themselves (on a 6-point scale; 1 = *Very much like me*, 6 = *Not like me at all*; we reverse-coded the scores for ease of comprehension). The two items in the respective dimensions were averaged given their moderate correlations.

Results

We conducted separate analyses for participants' averaged frequencies ($Mdn = 1.00$ meaning "*Never*") of own transgressions and perceived wrongness of others' transgressions ($M = 3.27$, $SD = 0.51$) due to their different data types and the highly skewed distribution of own transgressive frequencies.

We first performed a two-level linear mixed model, examining perceived wrongness of others' transgressions as a function of moral value (mean-centered; $M = 4.53$, $SD = 0.90$), reputation/status value (mean-centered; $M = 3.78$, $SD = 1.14$) and their two-way interaction, with individuals on the first level nested within countries on the second level (see Supplementary Table 1 for details). We found a positive effect of moral value ($B = 0.09$, $SE = 0.003$, $t = 32.47$, $p < .001$, 95% CI [0.082, 0.093]), but a negative effect of reputation/status value ($B = -0.03$, $SE = 0.003$, $t = -11.79$, $p < .001$, 95% CI [-0.038, -0.026]), on perceived wrongness of others' transgressions. Crucially, a significant two-way interaction ($B = 0.01$, $SE = 0.002$, $t = 3.05$, $p = .002$, 95% CI [0.006, 0.02]) revealed that moral value was associated with harsh judgment of others' transgressions, especially when people had stronger (+1 SD ; $B = 0.11$, $SE = 0.004$, $t = 24.98$, $p < .001$) rather than weaker (-1 SD ; $B = 0.08$, $SE = 0.003$, $t = 21.79$, $p < .001$) reputation/status value (see Figure 2).

We then fitted a two-level generalized linear mixed model on people's own frequency of transgressions, with moral value, reputation/status value and their interaction effect as predictors on the individual level and individuals nested within countries (see also Supplementary Table 1 for specifics). We found significant main effects of moral value ($B = -0.02$, $SE = 0.002$, $z = -6.94$, $p < .001$, 95%CI [-0.021, -0.012]) and reputation/status value ($B = 0.01$, $SE = 0.002$, $z = 4.53$, $p < .001$, 95%CI [0.006, 0.016]), but not their interaction ($B = -0.004$, $SE = 0.002$, $z = -1.86$, $p = .06$, 95%CI [-0.008, 0.001]; see also Figure 2). The findings suggested an association between strong moral value and fewer self-reported transgressions, irrespective of reputation/status value.

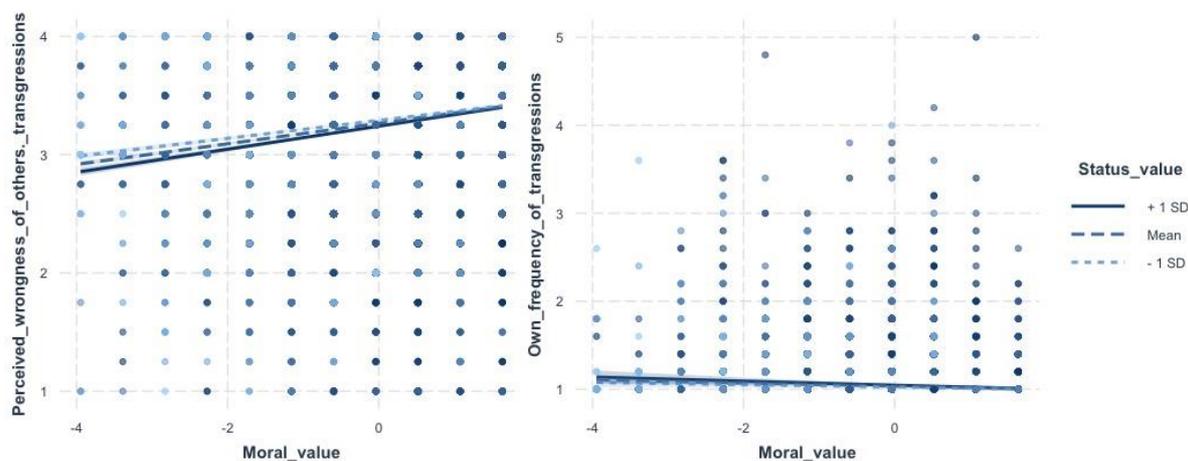


Figure 2. Perceived wrongness of others' economic transgressions (left panel) and frequency of own economic transgressions (right panel) as a function of moral value and reputation/status value in Study 2.

Discussion

In Study 2, moral character (as moral value) correlated with both harsher judgment of other wrongness and lower self-reported frequency of own transgressions. More importantly, consistent with Study 1, reputation management motives (as reputation/status value) exacerbated the positive link between moral character and harshness towards others' transgressions, but did not moderate the effects of moral character on the prevalence of own transgressions (i.e., moral hypocrisy).

Study 3

Although Study 2 examined real-life transgressions, people's own transgressions were measured with self-reports rather than actual observations. Moreover, transgressions in Study 2 were not perfectly aligned between measures referring to self and other. Therefore, in Study 3 we conducted a behavioral experiment where people evaluated identical (im)moral behaviors in the self and other conditions. In the self condition, the experimental task examined people's actual (un)fair choices in resource distribution, and their subsequent evaluations of their own choices.

In this last study, we captured moral character with one of its most representative measures in studies of moral judgment and behavior, that is, moral identity (Aquino & Reed, 2002; Hertz & Krettenauer, 2016). Moreover, as a measure of reputation management motives, we asked people to indicate their concern about status, which was found to influence fairness decisions in interactions with others (Blader & Chen, 2011, 2012).

Method

Participants. We predetermined to recruit 200 participants. A priori power analysis indicated a sample of $N = 126$, with 80% power to detect a medium-size (i.e., $\eta_p^2 = .06$ as in Study 1) effect of our intended three-way interaction in a linear regression at an alpha level of 0.05. Two hundred and one American participants (117 males; $M_{\text{age}} = 37.9$ years, $SD = 10.5$) completed our experiment through the crowdsourcing platform TurkPrime.com (Peer et al., 2017), and were all included in further analyses.

Design and procedure. Participants first indicated their moral character, operationalized as moral identity. The moral identity measure was the 5-item internalization subscale of the Self-Importance of Moral Identity Questionnaire (Aquino & Reed, 2002). Participants were asked to read nine adjectives (e.g., generous, fair) and rated self-importance of these characteristics (e.g., “It would make me feel good to have these characteristics”); on a 7-point scale ranging from 1 = *Strongly disagree* to 7 = *Strongly agree*; $\alpha = 0.73$).

Participants then completed a short survey ostensibly about their work status³. To measure reputation management motives, we adopted a validated measure of concern about status in interaction tasks (Blader & Chen, 2011, 2012; e.g., “I find it important that others acknowledge my status” “I wish to have high status”; $\alpha = 0.91$ across 10 items).

Participants were eventually enrolled in an online interaction game and were randomly assigned to one of two roles, described as either an Observer ($n = 96$) or a Distributor ($n = 105$). The online interaction game was an incentivized dictator game with

³ There was another 8-item scale measuring participants’ current status at work (Yu et al., 2019; e.g., “In general, my position tends to be highly respected”; $\alpha = 0.92$), which was not significantly associated with moral acceptability judgment ($t = 1.83$, $p = .07$), and was thus not included in the main analyses.

third-party punishment (see SM for detailed instructions). We first explained that Distributors should assign a \$10 bonus between the self and another online Recipient, with two options: (1) \$8 to the self and \$2 to the Recipient or (2) \$5 to the self and \$5 to the Recipient⁴. We then explained that Observers had a \$5 bonus and could choose to deduct the Distributors' payoff at a 1:3 ratio (that is, if the Observer pays \$0.1, the Distributor loses \$0.3 without influencing the Recipient's outcome). Participants were then randomly assigned to their role of a Distributor or an Observer. After making a Distributor choice, Distributors indicated "How acceptable do you think your decisions were in the game?" (on a 7-point scale ranging from -3 = *Completely unacceptable* to 3 = *Completely acceptable*). In the Observer role, participants reviewed all potential Distributor choices. For each potential choice, they indicated "How acceptable do you think the Distributor's decisions were in the game?" on the same scale, and indicated the amount they wanted to pay to reduce the Distributor's bonus (from \$0 to \$2.5 with unit increments of \$0.1). The Distributors and Observers were later matched into pairs, and 5% of them received the actual bonus as determined by their choices.

Results

Among the included Distributors ($N = 102$), 54 (52.9%) participants made an \$8/\$2 offer and 48 (47.1%) participants made a \$5/\$5 offer. We first analyzed Distributors' actual behavior, and then contrasted their judgment of own behavior with Observers' judgment of identical behavior.

⁴ An online dice rolling procedure (for specifics, see the SM) was also introduced to help Distributors choose between (1) \$8/\$2 and (2) \$5/\$5. However, very few Distributors ($n = 3$) chose to roll the dice. We therefore excluded this category from further analyses and mainly focused on participants who chose \$8/\$2 or \$5/\$5 directly.

Moral behavior. We conducted a binary logistic regression to examine the effects of mean-centered moral identity ($M = 5.59$, $SD = 1.11$) and status concern ($M = 3.80$, $SD = 1.63$), and their interaction, on Distributors' choice between \$8/\$2 ($= 0$) and \$5/\$5 ($= 1$). None of the effects of moral identity ($z = 1.46$, $p = .14$), status concern ($z = -1.81$, $p = .07$), or their interaction ($z = 0.40$, $p = .69$) were significant.

Moral judgment. We contrasted judgments of a \$8/\$2 offer in a linear regression, as a function of moral identity, status concern, and moral target (i.e., Observers judging others whilst Distributors evaluating the self). We found significant main effects of moral identity ($B = -0.37$, $SE = 0.17$, $t = -2.16$, $p = .03$, $\eta_p^2 = .03$, 95% CI [-0.71, -0.03]) and moral role ($B = -1.08$, $SE = 0.18$, $t = -5.85$, $p < .001$, $\eta_p^2 = .19$, 95% CI [-1.45, -0.72]). More importantly, a three-way interaction effect emerged ($B = -0.31$, $SE = 0.13$, $t = -2.40$, $p = .02$, $\eta_p^2 = .04$, 95% CI [-0.57, -0.05]; see Figure 3). When people were less concerned with status ($-1 SD$), their moral identity predicted lower acceptability judgments of their own (i.e., Distributor; $B = -0.95$, $SE = 0.47$, $t = -2.04$, $p = .04$) but not others' selfish choices (i.e., Observer; $B = 0.12$, $SE = 0.30$, $t = 0.42$, $p = .68$). In contrast, with high status concerns ($+1 SD$), people with stronger moral identity judged others' ($B = -0.64$, $SE = 0.25$, $t = -2.53$, $p = .01$), but not their own ($B < 0.01$, $SE = 0.33$, $t < 0.01$, $p > .99$), selfish choice as less acceptable. None of other predictors had a significant effect ($ps > .12$, $\eta_p^2 < .02$). The moral target by moral identity by status concern three-way interaction effect only emerged in judgments of a selfish \$8/\$2 but not a non-selfish \$5/\$5 offer ($B = -0.09$, $SE = 0.06$, $t = -1.54$, $p = .13$, $\eta_p^2 = .02$, 95% CI [-0.21, 0.03]).

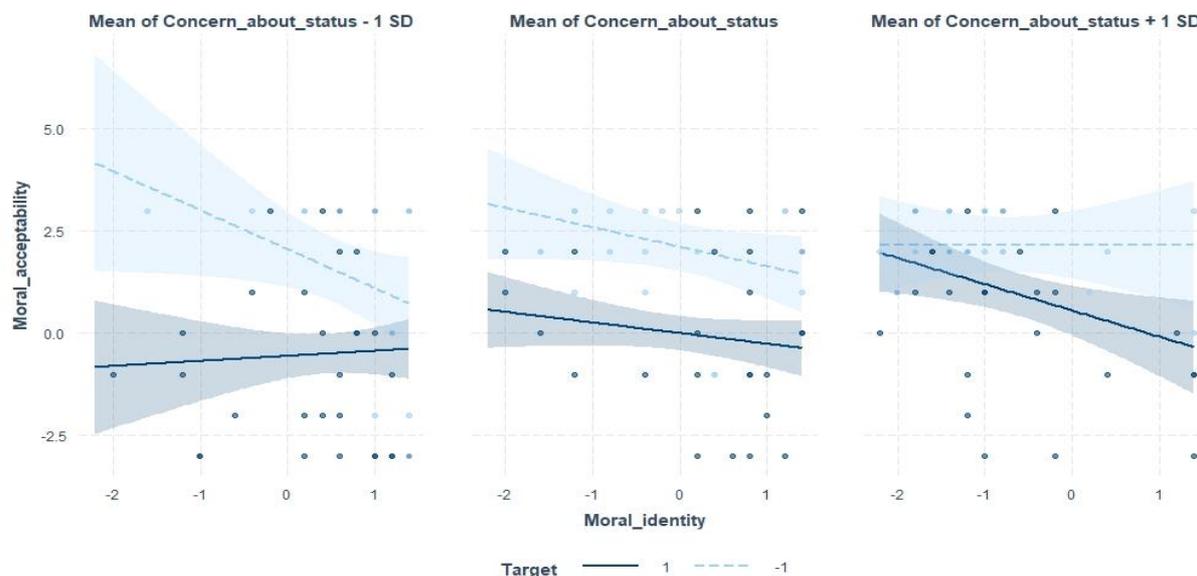


Figure 3. Moral acceptability judgments of own (= -1) and others' (= 1) selfishness in an actual dictator game, as a function of moral identity and concern about status in Study 3.

Discussion

In Study 3, moral character (as moral identity) related to lower acceptability judgments of selfish behavior, for both actors and observers. However, as in Studies 1 and 2, this effect was influenced by reputation management motives (as concern about status). When people were more (vs. less) concerned about their status, moral character predicted lower acceptability judgments of others' but not own selfishness (i.e., more moral hypocrisy).

General Discussion

A well-known Golden Rule of morality is to treat others as you wish to be treated yourself (Singer, 1963). People with a strong moral character might be expected to follow this Golden Rule, and judge others no more harshly than they judge themselves. But from an

evolutionary perspective, cultivating a strong moral character serves a reputation management function (Heintz et al., 2016; Trivers, 1971), and facilitates socially desirable reactions in a prompt and heuristic way (Hardy & Van Vugt, 2006; Jordan et al., 2016; Jordan & Rand, 2020). In this case, criticizing others' transgressions can be a relatively cost-effective approach to demonstrate one's moral character to others (Crockett, 2017; Jordan et al., 2017; Simpson et al., 2013). To the extent that the cultivation of a strong moral character is driven by reputation management motives, we therefore predicted that it would be related to increased hypocrisy, that is, moral harshness towards others more than towards the self. Across judgments of hypothetical scenarios in Study 1, reports of daily-life transgressions in Study 2, and evaluations of behavior in an incentivized experiment in Study 3, three studies consistently revealed among people with stronger (vs. weaker) reputation management motives, moral character predicted more hypocrisy, that is, harsher judgments of others' transgressions but not stricter standards for own misdeeds.

Previous theorizing generally implies moral character as genuinely and unconditionally good (Aquino & Reed, 2002; Kamtekar, 2004; Walker et al., 1987; Walker & Frimer, 2007); consistently, we found a positive correlation between moral character and stringent moral judgment across the three studies. More importantly, we complement these previous insights by emphasizing the strategic aspects of moral character in how people judge others versus themselves consistent with reputation management goals. People with a strong moral character can be sensitive to moral contexts, and tailor their moral performances accordingly. Thus, when motivated by reputation management, individuals with a strong

moral character also demonstrated more hypocrisy by displaying more moral harshness towards others than towards themselves. Such self-other discrepancies seem to be most likely to emerge when it serves a positive reputation. Correspondingly, as a function of strong moral character and reputation management motives, people only condemned others' selfishness more harshly but did not appraise others' fairness more positively (Study 3). A possible explanation for this finding is that particularly condemnation is effective to demonstrate one's righteousness to others (Baumeister et al., 2001; Jordan et al., 2017).

We employed diverse samples and methods to test the reputation management account of moral character; however, at least two important limitations should be noted, which merits future research. First, although our findings showed a positive relationship between moral character and hypocrisy, it is not yet clear whether people subjectively experience their judgments and behaviors as hypocritical. For example, people often feel morally superior to others (Guenther & Alicke, 2010), and therefore, moral identity can make people feel more entitled than others to commit certain transgressions (Dong et al., 2019). Second, we only measured self-reported moral character, which may not always align with perceived moral character as indicated by independent third parties. Indeed, such a discrepancy is likely, as individuals who self-claim a good moral character yet display hypocritical behaviors may be judged to have a worse moral character (Jordan et al., 2017).

To conclude, how moral character guides moral judgments and deeds depends on reputation management motives. When people are motivated to attain a good reputation and high status, their moral character may stimulate more hypocrisy by displaying stronger moral

harshness towards others than towards themselves. Thus, seeing oneself as a moral person does not always translate into doing good deeds, but can manifest as showcasing one's morality to others. Desires for a positive reputation might help illuminate why (some) moral people are susceptible to accusations of hypocrisy—for applying higher moral standards to others than to themselves.

References

- Althof, W., & Berkowitz, M. W. (2006). Moral education and character education: Their relationship and roles in citizenship education. *Journal of Moral Education, 35*(4), 495-518. <https://doi.org/10.1080/03057240601012204>
- Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin, 141*(3), 574–601. <https://doi.org/10.1037/a0038781>
- Aquino, K., & Reed, A. II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440. <https://doi.org/10.1037/0022-3514.83.6.1423>
- Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology, 72*(6), 1335-1348. <https://doi.org/10.1037/0022-3514.72.6.1335>
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology, 77*(3), 525-537. <https://doi.org/10.1037//0022-3514.77.3.525>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology, 5*(4), 323-370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Blader, S. L., & Chen, Y. R. (2011). What influences how higher-status people respond to lower-status others? Effects of procedural fairness, outcome favorability, and

concerns about status. *Organization Science*, 22(4), 1040-1060.

<https://doi.org/10.1287/orsc.1100.0558>

Blader, S. L., & Chen, Y.-R. (2012). Differentiating the effects of status and power: A justice perspective. *Journal of Personality and Social Psychology*, 102(5), 994–

1014. <https://doi.org/10.1037/a0026651>

Brick, C., Sherman, D. K., & Kim, H. S. (2017). “Green to be seen” and “brown to keep down”: Visibility moderates the effect of identity on pro-environmental behavior. *Journal of Environmental Psychology*, 51, 226-238.

<https://doi.org/10.1016/j.jenvp.2017.04.004>

Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character.

Current Directions in Psychological Science, 21(5), 355-359.

<https://doi.org/10.1177/0963721412454874>

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1(11),

769-771. <https://doi.org/10.1038/s41562-017-0213-3>

Dong, M., van Prooijen, J. W., & van Lange, P. A. (2019). Self-enhancement in moral

hypocrisy: Moral superiority and moral identity are about better appearances. *PloS*

One, 14(7), e0219382. <https://doi.org/10.1371/journal.pone.0219382>

ESS Round 2: European Social Survey Round 2 Data (2004). Data file edition 3.6. NSD -

Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS

data for ESS ERIC. <https://doi.org/10.21338/NSD-ESS2-2004>

Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences*, 36(5), 1015-1029.

[https://doi.org/10.1016/S0191-8869\(03\)00197-1](https://doi.org/10.1016/S0191-8869(03)00197-1)

Flynn, F. J., Reagans, R. E., Amanatullah, E. T., & Ames, D. R. (2006). Helping one's way to the top: Self-monitors achieve status by helping others and knowing who helps whom. *Journal of Personality and Social Psychology*, 91(6), 1123-1137.

<https://doi.org/10.1037/0022-3514.91.6.1123>

Graham, J., Meindl, P., Koleva, S., Iyer, R., & Johnson, K. M. (2015). When values and behavior conflict: Moral pluralism and intrapersonal moral hypocrisy. *Social and*

Personality Psychology Compass, 9(3), 158–170. <https://doi.org/10.1111/spc3.12158>

Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: Status, reputation, and conspicuous conservation. *Journal of Personality and Social*

Psychology, 98(3), 392–404. <https://doi.org/10.1037/a0017346>

Guenther, C. L., & Alicke, M. D. (2010). Deconstructing the better-than-average effect.

Journal of Personality and Social Psychology, 99(5), 755-770.

<https://doi.org/10.1037/a0020959>

Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, 32(10), 1402-1413.

<https://doi.org/10.1177/0146167206291006>

Heintz, C., Karabegovic, M., & Molnar, A. (2016). The co-evolution of honesty and strategic vigilance. *Frontiers in Psychology*, 7, 1503. <https://doi.org/10.3389/fpsyg.2016.01503>

Hertz, S. G., & Krettenauer, T. (2016). Does moral identity effectively predict moral behavior?: A meta-analysis. *Review of General Psychology, 20*(2), 129-140.

<https://doi.org/10.1037/gpr0000062>

Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences, 113*(31), 8658-8663. <https://doi.org/10.1073/pnas.1601280113>

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. <https://doi.org/10.1037/pspi0000186>

Jordan, J. J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science, 28*(3), 356-368. <https://doi.org/10.1177/0956797616685771>

Kamtekar, R. (2004). Situationism and virtue ethics on the content of our character. *Ethics, 114*(3), 458-491. <https://doi.org/10.1086/381696>

Kohlberg, L. (1976). Moral stages and moralization: The cognitive developmental approach. In T. Lickona (Ed.), *Moral development and behavior: Theory, research, and social issues* (pp. 31-53). New York: Holt, Rinehart & Winston.

Lammers, J. (2012). Abstraction increases hypocrisy. *Journal of Experimental Social Psychology, 48*(2), 475-480. <https://doi.org/10.1016/j.jesp.2011.07.006>

Lönnqvist, J. E., Rilke, R. M., & Walkowitz, G. (2015). On why hypocrisy thrives:

Reasonable doubt created by moral posturing can deter punishment. *Journal of Experimental Social Psychology*, *59*, 139-145.

<https://doi.org/10.1016/j.jesp.2015.04.005>

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063),

1291-1298. <https://doi.org/10.1038/nature04131>

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of*

Personality and Social Psychology, *46*(3), 598–609. [https://doi.org/10.1037/0022-](https://doi.org/10.1037/0022-3514.46.3.598)

[3514.46.3.598](https://doi.org/10.1037/0022-3514.46.3.598)

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative

platforms for crowdsourcing behavioral research. *Journal of Experimental Social*

Psychology, *70*, 153-163. <https://doi.org/10.1016/j.jesp.2017.01.006>

Polman, E., & Ruttan, R. L. (2012). Effects of anger, guilt, and envy on moral

hypocrisy. *Personality and Social Psychology Bulletin*, *38*(1), 129–139.

<https://doi.org/10.1177/0146167211422365>

Reed, A. II, & Aquino, K. F. (2003). Moral identity and the expanding circle of moral regard

toward out-groups. *Journal of Personality and Social Psychology*, *84*(6), 1270-

1286. <https://doi.org/10.1037/0022-3514.84.6.1270>

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social

psychology quantitatively described. *Review of General Psychology*, *7*(4), 331-363.

<https://doi.org/10.1037/1089-2680.7.4.331>

- Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity. *European Journal of Psychological Assessment*, 21(3), 202-211. <https://doi.org/10.1027/1015-5759.21.3.202>
- Shaw, A., Montinari, N., Piovesan, M., Olson, K. R., Gino, F., & Norton, M. I. (2014). Children develop a veil of fairness. *Journal of Experimental Psychology: General*, 143(1), 363–375. <https://doi.org/10.1037/a0031247>
- Simpson, B., Harrell, A., & Willer, R. (2013). Hidden paths from morality to cooperation: Moral judgments promote trust and trustworthiness. *Social Forces*, 91(4), 1529-1548. <https://doi.org/10.1093/sf/sot015>
- Singer, M. (1963). The Golden Rule. *Philosophy* Vol. XXXVIII #146: 293-314.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30(4), 526–537. <https://doi.org/10.1037/h0037039>
- Sperber, D., & Baumard, N. (2012). Moral reputation: An evolutionary and cognitive perspective. *Mind & Language*, 27(5), 495-518. <https://doi.org/10.1111/mila.12000>
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35-57. <https://doi.org/10.1086/406755>
- Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44(5), 1334-1338. <https://doi.org/10.1016/j.jesp.2008.03.010>

- Vonasch, A. J., Reynolds, T., Winegard, B. M., & Baumeister, R. F. (2018). Death before dishonor: Incurring costs to protect moral reputation. *Social Psychological and Personality Science*, 9(5), 604-613. <https://doi.org/10.1177/1948550617720271>
- Walker, L. J., de Vries, B., & Trevethan, S. D. (1987). Moral stages and moral orientations in real-life and hypothetical dilemmas. *Child Development*, 842-858.
- Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology*, 93(5), 845-860. <https://doi.org/10.1037/0022-3514.93.5.845>
- Weiss, A., Burgmer, P., & Mussweiler, T. (2018). Two-faced morality: Distrust promotes divergent moral standards for the self versus others. *Personality and Social Psychology Bulletin*, 44(12), 1712-1724. <https://doi.org/10.1177/0146167218775693>
- Yoder, K. J., & Decety, J. (2014). The good, the bad, and the just: Justice sensitivity predicts neural response during moral evaluation of actions performed by others. *Journal of Neuroscience*, 34(12), 4161-4166. <https://doi.org/10.1523/JNEUROSCI.4648-13.2014>
- Yu, A., Hays, N. A., & Zhao, E. Y. (2019). Development of a bipartite measure of social hierarchy: The perceived power and perceived status scales. *Organizational Behavior and Human Decision Processes*, 152, 84-104. <https://doi.org/10.1016/j.obhdp.2019.03.011>