Moral Goodness is Gendered

Code and Supplemental materials:
https://osf.io/7p6e9/?view_only=f4f9310683db48e39240d0fcc215ebcd

Abstract

Women are consistently judged to be more caring and communal than men. We find that moral goodness itself is feminine at the level of faces, traits, behaviors, and impressions (N = 5,376). Using reverse-correlation methods, we show that mental representations of the face of "a morally good person" is a woman's face. When completing a novel profile-generation task, people ascribe more morally good traits to women than to men, both when they spontaneously think of a woman (vs. a man) and when assigned. Across 1,864 previously normed behaviors, we find that actions associated with moral goodness are stereotyped as feminine. Finally, in a novel impression updating task, we find that immoral behavior is more diagnostic for women than men; equivalent moral transgressions lead to larger negative character updating for women than for men. In sum, moral goodness is gendered: people think women are morally superior to men.

Keywords: morality, gender, moral goodness, moral constraints

Research Transparency Statement

General Disclosures
Conflicts of interest: We have no conflicts of interest to report. Funding: Blinded. Ethics: This research received approval from local ethics boards (IDs: Blinded).

Experiment 1
Preregistration: The hypotheses, analyses, and methods for the ratings study (part 2 of Experiment 1) were preregistered on November 22nd, 2024, prior to data collection which began later that day on November 22nd, 2024. There was a minor deviation (a change from using a matched t-test to using a one-sample t-test) from the preregistered analysis plan, which we note in the main text of the results section. Materials: All study materials, data, the preregistration, and analysis scripts are publicly available (https://osf.io/7p6e9/?view_only=f4f9310683db48e39240d0fcc215ebcd).

Experiment 2
Preregistration: Experiment 2A was not preregistered. Experiment 2B was a preregistered replication of Experiment 2A. The hypotheses, analyses, and methods for this experiment were preregistered on December 26th, 2024, prior to data collection which began later in the day on December 26th, 2024. We note any deviations from the preregistered analysis plan in the main text, including removing 2 additional participants (outside of the preregistered exclusion criteria) for having duplicate IP addresses. Materials: All study materials, data, the preregistration, and analysis scripts are publicly available (https://osf.io/7p6e9/?view_only=f4f9310683db48e39240d0fcc215ebcd).

Experiment 3
Preregistration: Experiment 3 was not preregistered. We sought to test the hypothesis in a previously collected dataset. Materials: All study materials, data, and analysis scripts are publicly available (https://osf.io/7p6e9/?view_only=f4f9310683db48e39240d0fcc215ebcd).

Experiment 4
Preregistration: The hypotheses, analyses, and methods were added to the preregistration on December 30th, 2024, prior to data collection which began later in the day on December 30th, 2024. Materials: All study materials, data, the preregistration, and analysis scripts are publicly available (https://osf.io/7p6e9/?view_only=f4f9310683db48e39240d0fcc215ebcd).

Moral Goodness is Gendered

Are women wonderful? People seem to think so. In 1994, Alice Eagly and Antonio Mladinic challenged the then-accepted view that men are seen more positively than women. They articulated a puzzle: despite apparent societal disadvantages of women, people actually have more positive views of them—something they coined as "the women-are-wonderful effect" (see also Williams & Best, 1990). Subsequent work found that this effect holds using multiple measures, including implicit ones (Leach et al., 2017; Rudman & Goodwin, 2004), and persists in terms of both what people think women are and how they expect them to be (Prentice & Carranza, 2002), though it is somewhat attenuated in more gender-egalitarian societies (e.g., Denmark, Norway; Krys et al., 2017).

These positive evaluations of women are grounded in the perceived communal nature of women (e.g., Eagly & Karau, 1991; Flores-Robles & Gantman, 2024; Glick & Fiske, 2001; Heilman, 1983, Rudman & Phelan, 2008; see Hsu et al., 2021 for review). People believe that women are innately more sensitive, nurturing, pure, and less aggressive than men (Glick & Fiske, 2001), and that they are naturally suited for caretaker roles, both in the home and the workplace (e.g., Cejka & Eagly, 1999; Heilman, 2012; Hideg & Ferris, 2016; Prentice & Carranza, 2002; Rudman & Phelan, 2008). Men, on the other hand, are perceived as agentic—assertive, decisive, and in control (see Hsu et al., 2021 for review). The communal/agentic dichotomy is both gendered (Martin & Slepian, 2021) and pervasive, emerging consistently across time and corpora (Charlesworth et al., 2021) including letters of recommendation (Madera et al., 2009) and obituaries (Markowitz et al., 2025). This dichotomy is the cornerstone of benevolent sexism, or the set of beliefs that women, unlike men, are pure and good, and in need of protection, which are endorsed to varying degrees as an individual difference (Glick & Fiske, 2001).

How does this bear on moral judgment? Caring and communal traits are characteristic of morally good behavior (e.g., Graham et al., 2013; Gray et al., 2014; Fiske, 2018; Rai & Fiske, 2011; Janoff-Bulman & Carnes, 2013). While theories differ on the core constructs that constitute morality (e.g., Graham et al., 2013; Gray et al., 2014; Curry, 2016), principles, and obligations regarding harm and care are consistent among them. Moral traits are also conceptually distinct from other positive and stereotypically feminine traits like warmth or sociability (e.g., Brambilla et al., 2011; 2019)—and primary in our social evaluations of others (Goodwin et al., 2014).

Though women's relative goodness and purity are part of the constellation of beliefs that constitute benevolent sexism (e.g., Glick & Fisk, 2001), our work shifts focus to ask whether morality as a concept is feminized. We take general moral goodness as the starting construct and use a variety of methods—from mental representations (Study 1) to questions about traits (Study 2a and 2b), behaviors (Study 3), and character updating (Study 4)—to test not only whether women are associated with moral goodness, but also whether moral goodness is feminized. All research was approved by the [BLINDED] IRBs, and all related data, preregistrations (Study 1, 2b, and 4), materials, and code can be accessed on the project's Open Science Framework page[1].

Study 1

Study 1 used a two-phase procedure to test whether mental representations of moral goodness are feminine. We used a reverse-correlation study—a data driven approach that allows researchers to visualize internal representations of concepts (Dotsch et al., 2008; Dotsch & Todorov, 2012)—to generate people's representations of the faces of morally good or morally bad people. In phase two, participants rated the average images produced from phase one on relevant

[1] https://osf.io/7p6e9/?view_only=f4f9310683db48e39240d0fcc215ebcd

dimensions. We hypothesized that the image from the morally good condition would be rated as more feminine than masculine, whereas the image from the morally bad condition would be rated more masculine than feminine.

Method

Sample size and power analysis

We recruited a total of 292 participants balanced on gender from Prolific to complete the initial reverse-correlation task. We excluded 22 participants for failing comprehension questions or self-reporting that they did not follow attention for a total analyzable sample of 270 (47% Men, 50% Women, 3% Non-binary, Other, or missing; M_{age}: 36.62, SD_{age}: 12.03). Once reverse-correlated images were generated for each of the two conditions, a separate sample of participants (N = 220) balanced on gender was recruited from Prolific to rate the average images. We conducted an a priori power analysis using G*Power (Faul et al., 2009) to determine the sample size needed for the image ratings. To have 80% power to detect a small effect (d = 0.20), a total of 199 participants was required. We also preregistered removing any participants who failed the bot or attention check item (N = 10). We also removed participants who had duplicate IP addresses (N = 4), leaving us with a total sample of 212 (48.1% Women, 51.4% Men, 0.47% non-binary; Race: 58.96% White, 25.00% Black, 16.04% Other 3%; Mean age: 37.39, SD age: 13.29). All participants then reported basic demographic information and were debriefed and compensated for their time

Stimuli Creation

The base face image for the reverse correlation task was an average of 60 Chicago Face Database (Ma et al., 2015) matched on race (all white), attractiveness, and trustworthiness across gender (30 male, 30 female) computed using photoshop (see Figure 1). The base image was then used to create the 350 task stimuli superimposed with random noise patterns using the default settings from the “rcicr” package (Dotsch, 2023). When random noise is added to the base image, the noise creates different highlights and shadows on the face, producing a series of stimuli that look visually distinct even though the base image is constant. See Supplemental Materials for more information on the stimuli and classification images.

Procedure and Design

Participants first completed the reverse-correlation task, classifying faces as belonging to either a morally good or bad person, depending on condition (Phase One). Next, a separate sample of participants evaluated the images produced from the reverse-correlation task on how masculine or feminine they appeared, as well as a series of additional traits related to moral goodness and agency (Phase Two).

Phase One

The procedure of the reverse correlation task mirrored that of prior work (e.g., Brown-Iannuzzi et al., 2017). A total of 270 participants completed the first portion of the experiment. Each participant was randomly assigned to one of the two between-subjects moral valence conditions: morally good or morally bad. The experiment was displayed using *JsPsych* (De Leeuw et al., 2015). Participants were shown two side-by-side face images and selected which of the two looked more morally good or bad (depending on condition) for 350 pairs of images (see Figure 1A for schematic). We then created averages of the participant-level classification images into an overall classification image separately for the morally good condition and one for the morally bad condition (Figure 1B; see supplemental material for additional information).

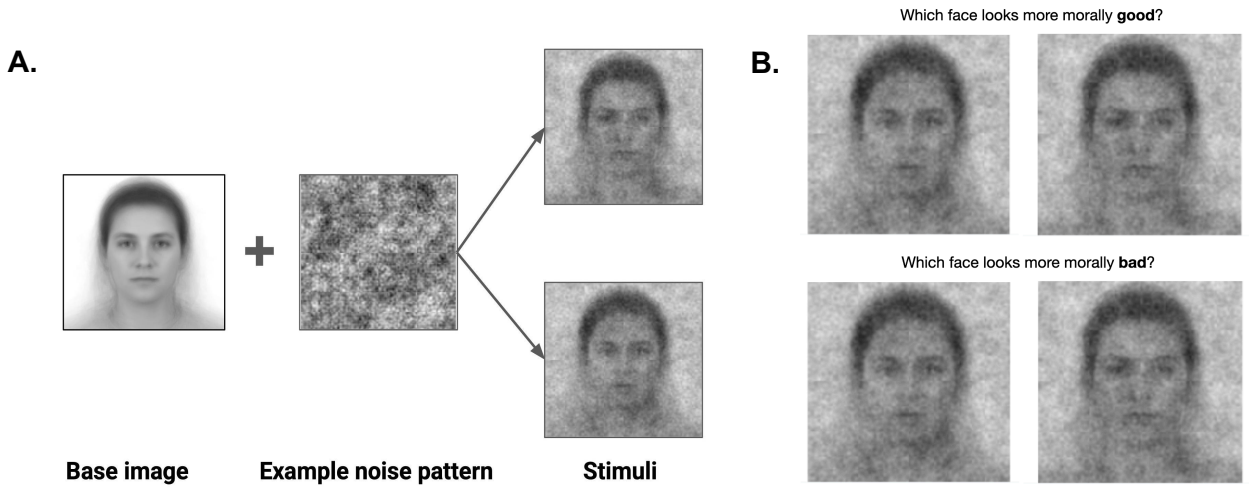


Figure 1. Panel A: Example of the base image used in the reverse-correlation task and examples of the stimuli presented to participants. Panel B: Example trial from each condition (between-subjects).

Phase Two

Next, we sought to test the association of morality with male/female faces in a pre-registered study. An independent sample of 212 participants completed this portion of the study. Participants rated the average classification image from the morally bad condition as well as the average classification image from the morally good condition (in random order) on a series of traits. Participants rated the images on the primary dimension of interest, gender (1 = *Very masculine* to 7 = *Very feminine*) and additionally completed a gender categorization measure, where participants were told to guess whether each image (separately) was more likely to be a man or a woman. Participants also rated the moral valence (1 = *Extremely morally* bad to 8 = *Extremely morally good*) of the image as a manipulation check. We predicted that participants would rate the classification image from the morally good condition as more feminine and more morally good. We additionally included a measure of the freedom to break rules and competence as measures of agency for the individual classification images. Those analyses are reported in the Supplemental Material.

After rating both images, participants completed an additional rating task assessing links between moral goodness, moral traits, positive attributes, and agency. In this task, the two classification images served as the anchors on a scale (with the morally bad person on the left and morally good on the right). We asked people to move the anchor from the middle to the face that better represented each of the following traits (all on 101-point scales): fair, caring, timid, wealthy, powerful, unconventional, compliant, creative, and funny. Creativity and humor/funny were included for exploratory purposes.

Results

All analyses in Study 1 were done by fitting a linear mixed-effects model with participants included as a random intercept and type of classification image (good vs bad) as the predictor as preregistered, or by using a one-sample t-test for each of the two images (*mu* = 50). All statistical tests reported in this manuscript are two-sided tests.

We first tested whether the classification image for moral goodness indeed appeared more morally good. As predicted, results suggested that participants saw the classification image from the morally good condition as more morally good (b = 1.30, SE = 0.11, t = 12.04, p < .001, *95%*

CI [1.09, 1.52], β = 1.00) than the classification image from the morally bad condition. Next, we tested our primary hypothesis. Critically, participants rated the morally good image as more feminine (*b* = 3.22, *SE* = 0.13, *t* = 24.29, *p* < .001, *95%CI* [2.96, 3.48], β = 1.47) than the morally bad one[2].

Next, we examined the results of the additional ratings task. We used one-sample t-tests (with a mean of 50) to test which image appeared to possess more of each trait. Although we did not preregister this approach, we determined during analysis that it was superior to mixed-effects models for testing these preregistered hypotheses derived from prior work (e.g., Eagly, 1987; Glick & Fiske, 2001; Heilman, 1983; Ellemers, 2018; Hsu et al., 2021). More specifically, we predicted that traits associated with moral goodness (fairness and caring) and traits associated with low agency (high timidity, low wealth, low power, conventional, and high compliance) would align more closely with a face perceived as morally good. Our hypotheses were supported: Participants rated the morally good image as being more reflective of care ($t(423) = 33.97$, $p < .001$, $d = 1.65$) and fairness ($t(423) = 19.93$, $p < .001$, $d = 0.97$). In contrast, the morally good image was seen as less reflective of agentic traits than the morally bad one: Participants rated the morally good image as lower in power ($t(423) = -7.84$, $p < .001$, $d = -0.38$), being less unconventional ($t(423) = -14.58$, $p < .001$, $d = -.71$), more compliant ($t(423) = 19.38$, $p < .001$, $d = 0.94$), and more timid ($t(423) = 4.30$, $p < .001$, $d = 0.21$) than the morally bad image. Contrary to predictions, participants rated the morally good image as more wealthy ($t(423) = 12.06$, $p < .001$, $d = 0.58$), which did not support predictions, but may reflect a positivity effect. See Figure 2 for results.

[2] We also included an exploratory categorization measure where participants were asked to predict the gender of each face. Participants overwhelmingly classified the morally good image as a woman—99% of participants guessed that the gender of the person in the image was a woman (vs. 60% categorizing the morally bad classification image as male; see Supplemental Materials).

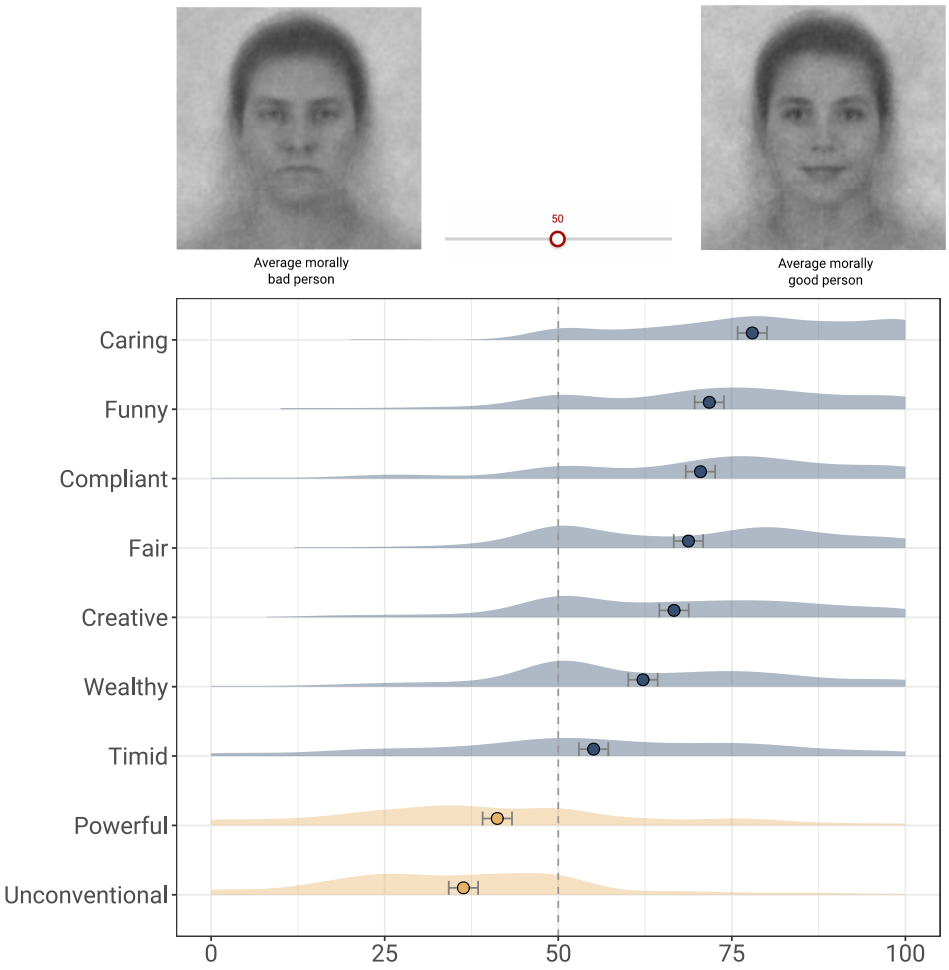


Figure 2. Mean ratings of traits where the two classification images served as anchors on the scale where the starting point on all ratings was the midpoint on the scale (50). Error bars represent 95% confidence intervals.

Discussion

Participants' representations of the face of moral goodness is a woman's face. Nearly all participants categorized the morally good image as being that of a woman, and traits associated with greater positivity (e.g., wealth) and moral goodness (e.g., fairness), but lower agency (e.g., compliance) were more aligned with the morally good image than the bad one. The morally bad classification image, in contrast, was seen as more powerful and free—hallmarks of agency (see e.g., Abele and Wojciszke, 2014).

These findings also extend the literature on face perception and gender stereotypes beyond associations with trustworthiness in particular (see Oosterhof & Todorov, 2008; Sutherland et al., 2015; Todorov & Oh, 2021) to morality in general. While trustworthiness and moral judgments tend to be correlated (e.g., Everett et al., 2016), they are distinct concepts/traits (e.g., a whistleblower: not trustworthy but just; evil henchman: trustworthy but immoral).

Study 2a

In Studies 2a and 2b, we tested whether specific moral traits (as well as morality overall) are more associated with women than with men both spontaneously (2a) and when prompted (2b). We

used a novel profile-generating paradigm where participants indicated how much of a series of traits to allocate to various target profiles. In Study 2a, participants allocated traits to profiles and were then asked whether they had a woman or a man in mind when they did so.

Method

Sample size and power analysis

We recruited a total of 375 individuals from Connect. We excluded participants who failed the attention check and bot check items (N = 22) or did not provide task data (N = 71), leaving us with a total sample of 282 (N = 282, 53% Women, 46% Men, ~1% Other or missing; M_{age} = 38.05, SD_{age} = 11.2) consenting participants. We ran a sensitivity analysis using G*Power (Faul et al., 2009). We specified a two-tailed test with alpha set to .05, and power set to 80%. This suggested that we were powered to detect a correlation of r = .16. The only inclusion criterion was current United States residence.

Avatar gender selection

To test whether gender influences ascriptions of moral traits, we prompted participants to indicate which gender (woman or man) they had in mind while completing each profile. Avatars for gender assessments were cartoons with purple skinned and had gray hair to remove social identity information not relevant to the current research question.

Procedure & Design

Participants were asked to make profiles of different people by allocating different magnitudes of eleven traits pulled from prior research (e.g., Graham et al., 2009; 2011; 2013; Curry, 2016; Schein, & Gray, 2018). Participants completed six total profiles (see Table 1) each consisting of 11 trait ratings presented in random order. Participants always rated themselves first followed by the other profiles in random order. All traits (see Table 1) were bipolar traits rated on 201-point scales with anchors beginning at the midpoint of the scale (see Supplemental Figure S4 for example). Most importantly, participants made a gender assessment immediately following their trait allocations for each profile except the one for themselves. Specifically, they selected which of two avatars—a man or a woman—better represented who they had in mind while making the ratings. Finally, all participants completed individual difference measures[3], provided demographic information, were debriefed, and were paid for their time.

Table 1.
Profiles and traits used in Study 2a.

Profiles	***Traits***
Yourself	Impartial - Loyal
An average person	Abstinent - Sexually Adventurous
An interesting leader	Values others first - Values self first
An interesting… fictional character	Straightforward - Artful
…person to be close to	Gentle - Rough
…person to meet one time	Cautious - Bold
	Equal opportunities - Helps friends
	Peaceful - Rebellious
	Angelic - Devilish
	Follow rules - break rules
	Civil - Challenging

[3] After all profiles were completed, participants completed a series of individual difference measures collected for a different research project and not analyzed here.

Results

Both Study 2a and 2b used the same analytic approach. All analyses were conducted using R statistical analysis software (R Core Team, 2013) using the "lme4" package (Bates et al., 2015), "lmerTest" package (Kuznetsova et al., 2017), and the "jtools" package (Long, 2022). P-values were calculated using Satterthwaite degrees of freedom approximation. Comparisons among individual profiles are reported in the Supplemental Material (e.g., excluding gender comparisons which are reported in the main text, conceptual mapping). Study 2a was exploratory.

To examine the effect of gender, we specified models that included the selected gender for each profile (or participant self-reported gender in the case of the "you" profiles), the profile target (e.g., average person, leader), and their interaction term as predictors (mean differences for each trait are reported in Table 1). Models included by-participant random slopes. We predicted that when people had a woman (vs. a man) in mind, they would assign more morally good traits to the profile.

As predicted, we find evidence that which gender people had in mind was associated with their moral trait ratings across profile targets (see Figure 3). Relative to women targets, participants rated men targets as more rough than gentle (b = 13.81, SE = 6.57, t = 2.10, *95%CI* [0.93, 26.69], p = .036, β = 0.23), more challenging than civil (b = 14.23, SE = 7.03, t = 2.03, *95%CI* [0.45, 28.02], p = .043, β = 0.23), more devilish than angelic (b = 12.36, SE = 6.26, t = 1.97, *95%CI* [0.07, 24.65], p = .049, β = 0.21), and more likely to break rules than follow them (b = 17.58, SE = 6.64, t = 2.65, *95%CI* [4.57, 30.60], p = .008, β = 0.28). There were also significant interactions for the "giving equal opportunities"/"helping friends" trait pair: Participants rated the targets as less helpful to friends (i.e., providing more equal opportunities) when the target they had in mind was a male character (b = -21.84, SE = 9.82, t = -2.22, *95%CI* [-41.10, -2.57], p = .026, β = 0.22) or the participant themselves was a man ("you" target; b = -28.35, SE = 9.75, t = -2.91, *95%CI* [-47.47, -9.22], p = .004, β = -0.45) compared to when the targets were women. Overall, when people spontaneously had women in mind when they created profiles for themselves, an average person, a leader, a person to be close to, or a person to meet one time, they were more likely to select more morally positive and more compliant traits.

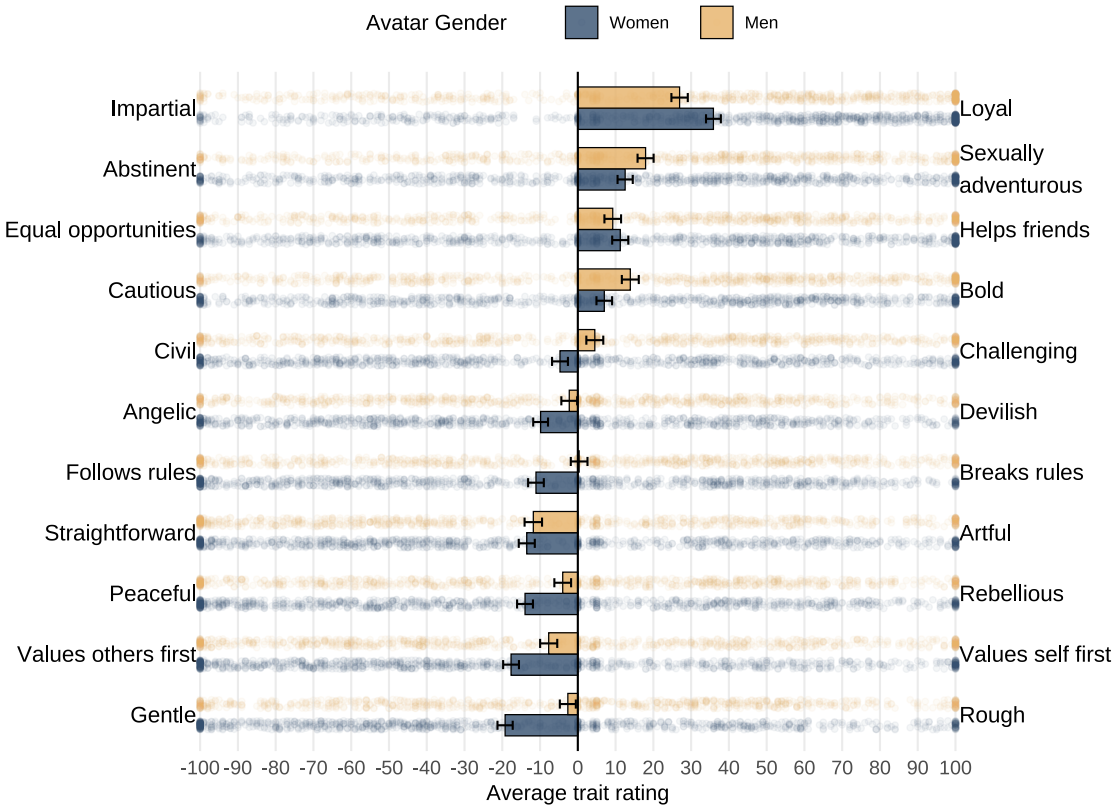


Figure 3. Mean ratings for each trait averaged across avatar gender for each profile. Datapoints represent individual participant data. Bars pointing in the negative direction indicate that the mean trait rating more closely aligns with traits listed on the left; positive values indicate alignment with traits on the right. Error bars represent the standard error of the mean.

Discussion

Individual moral traits are gendered. Profiles of men were rated lower on positive moral traits than those of women. Male targets were seen as more rough, challenging, devilish, and rule-breaking (traits often ascribed to evil targets; see Supplemental Material). However, it is possible that participants did not have a gender in mind until we prompted them, making their gender judgments post-hoc. To test whether people will deliberately assign men and women different degrees of moral traits, we pre-registered and conducted Study 2b.

Study 2b

In Study 2b, we preregistered, replicated, and extended Study 2a by manipulating the gender of the target profiles. We directly tested the causal role of gender in moral trait ascriptions— including traits associated with moral reasoning (e.g., fairness), which has been historically associated with men over women (Kohlberg, 1976). We predicted that average women's profiles would be seen as more morally good and closer to an ideal person (see also Bear & Knobe, 2017) than men's, even in fiction, where we thought women might be afforded more room to deviate from stereotypes. We also predicted there would be no difference in attributions of non-moral traits between average men and women profiles (despite many of these traits being positive). We also had several predictions about individual traits: We predicted the average woman would be rated as more generally moral, more compliant, less powerful, less willful, more conventional, and less assertive than the average man.

Method

Sample size and power analysis

We recruited a total of 500 people from Connect balanced on gender. We conducted an *a priori* power analysis using G*Power (Faul et al., 2009) to achieve 80% power to detect a small correlation effect (r = .18; for testing for correlations among different profiles[4]), which suggested 237 participants were necessary. We sought to recruit 250 women and 250 men. We preregistered removing participants who failed attention check or bot check items (N = 6). We additionally had two participants with duplicate IP Address information (N = 2) and removed the second instance in both cases. This left us with a final sample of 492 participants (N = 492; Gender: 50% women, 49% men, 1% non-binary; Mean age = 40.5, SD = 12.9).

Profiles

Participants completed five total profiles (see Figure 4) in the following order: the ideal person and the average person (presented in random order across participants), the average woman and man (random order across participants; see Figure 4 for example), and then an interesting fictional woman and man (random order across participants).

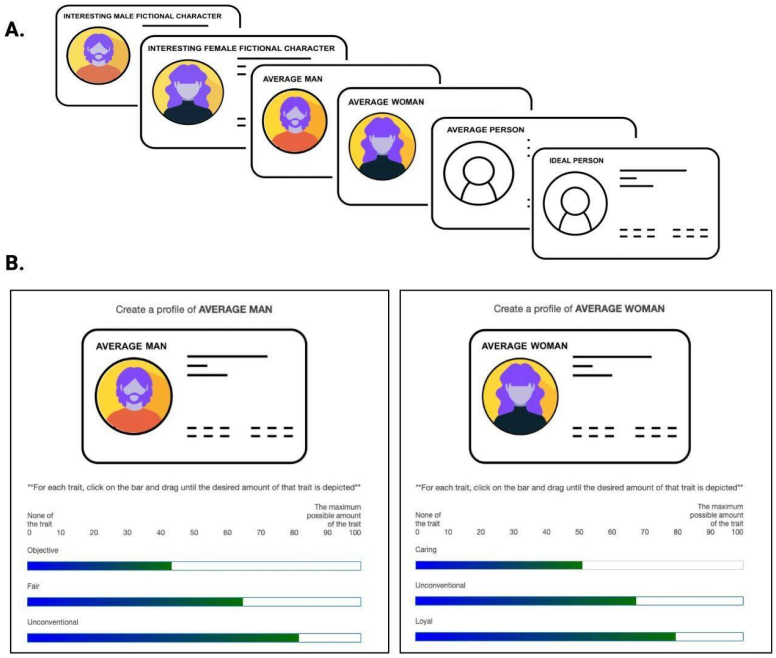


Figure 4. Panel A: Sequence of profile trials beginning with the ideal profile. Gender was presented in random order for profiles where it was manipulated. Panel B: Example of the average man and woman profiles and a random subset of traits rated by participants.

Procedure & Design

Profiles consisted of 17 total traits, including nine moral traits selected by drawing on numerous theories in morality (Abele & Wojciszke, 2007; Graham et al., 2013; Gray et al., 2014; Curry, 2016; Sun & Goodwin, 2020; Jackson, et al., 2023; Schein & Gray, 2018; see also Fiske, 2018), three non-moral traits, and five items related to rules and agency (e.g., Abele & Wojciszke, 2007) all rated on a 101-point scale where 0 = *none of the trait* and 100 = *maximum amount of the trait*. See Table 2 for individual traits. We preregistered collapsing the moral traits into a single

[4] We also conducted a sensitivity power analysis for comparisons between the average man and average woman. With our sample size and alpha set to .05, we had 80% power to detect d = .25.

composite index of morality (Cronbach's α = .95) and the nonmoral traits into a non-moral composite (Cronbach's α = .77).

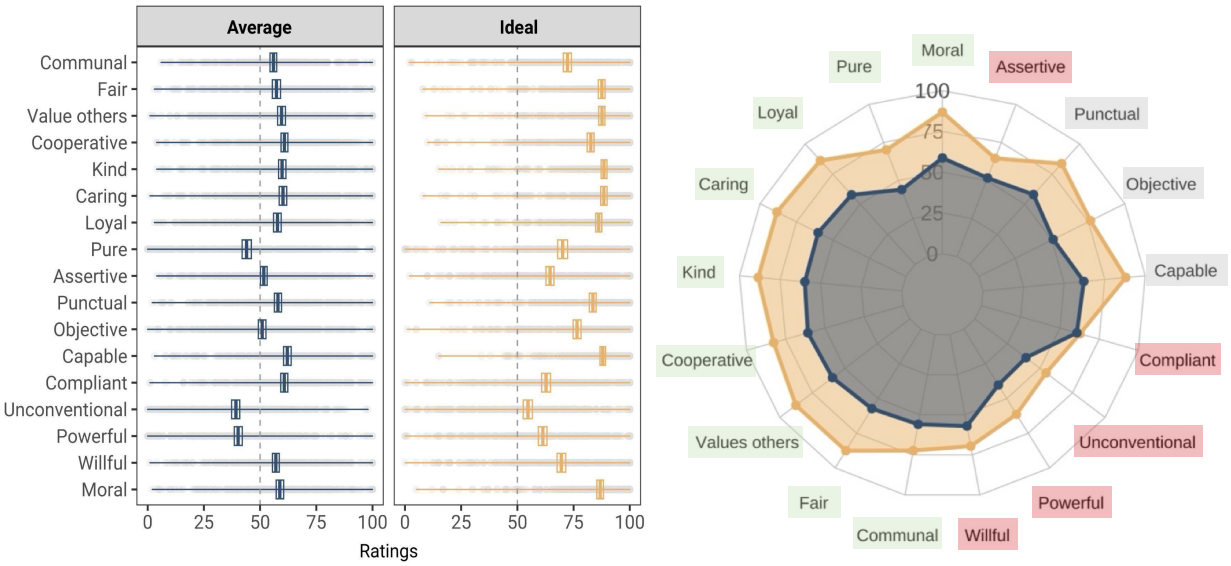
Table 2.
Traits selected for profile ratings in Study 2b.

Moral Traits	***Non-Moral Traits***	***Agency & Rules Traits***
Caring	Capable	Assertive
Valuing others	Objective	Compliant
Kindness	Punctual	Powerful
Purity		Unconventional
Fairness		Willful
Cooperativeness		
Loyalty		
Communality		
(General) Morality		

Note. Additional details on the selection of traits are reported in the Supplemental Material.

Results

First, as a manipulation check, we examined whether the ideal profile was rated as more moral than the average profile, using a linear mixed-effects model with profile target included as the predictor and by-participant random intercepts[5]. Looking at only trials that included the average and ideal profiles, we indeed find evidence in favor of our predictions. Ideal people are rated as more morally good than average ones, b = 26.27, SE = 0.78, t = 33.69, $95\%CI$ [24.74, 27.80], p < .001, d = 1.52 (see also Figure 5).



[5] We preregistered including by-target random intercepts as well as by-participant random slopes, however models did not converge. As preregistered, we first modified the optimizer and then simplified the random effects structure. We additionally include Cohen's d estimates based on raw difference scores and not extracted from the *lmer* model where possible.

Figure 5. Left Panel: Average individual trait ratings for the Average and Ideal profiles. Thick middle lines represent the mean. Edges of the box represent bootstrapped 95% CIs. Right Panel: Radar plot comparing Average profile ratings (in blue) to Ideal profile ratings (in yellow). Green labels represent moral traits, red labels represent agency traits of interest, and grey represent non-moral traits.

Critically, we found support for our primary hypothesis. The profiles of average women were rated as more morally good than average men, b = 11.54, SE = 0.68, t = 16.93, $95\%CI$ [10.20, 12.88], p < .001, d = 0.76. Similarly, when we instead use the individual moral traits to compare ratings of average men and women, we find consistent results—women are rated as more morally good across each individual trait (see Supplemental Material). We also predicted that profiles of fictional women would be rated as more morally good than fictional men, but found no such evidence, p = .14 (see Figure 6). Further, we predicted there would be no differences in how the composite nonmoral traits were ascribed across the gender of the average profiles. We do not find evidence of significant differences in the effect of gender (p = .067, d = 0.06), but also do not find evidence of equivalence (see Supplemental Material).

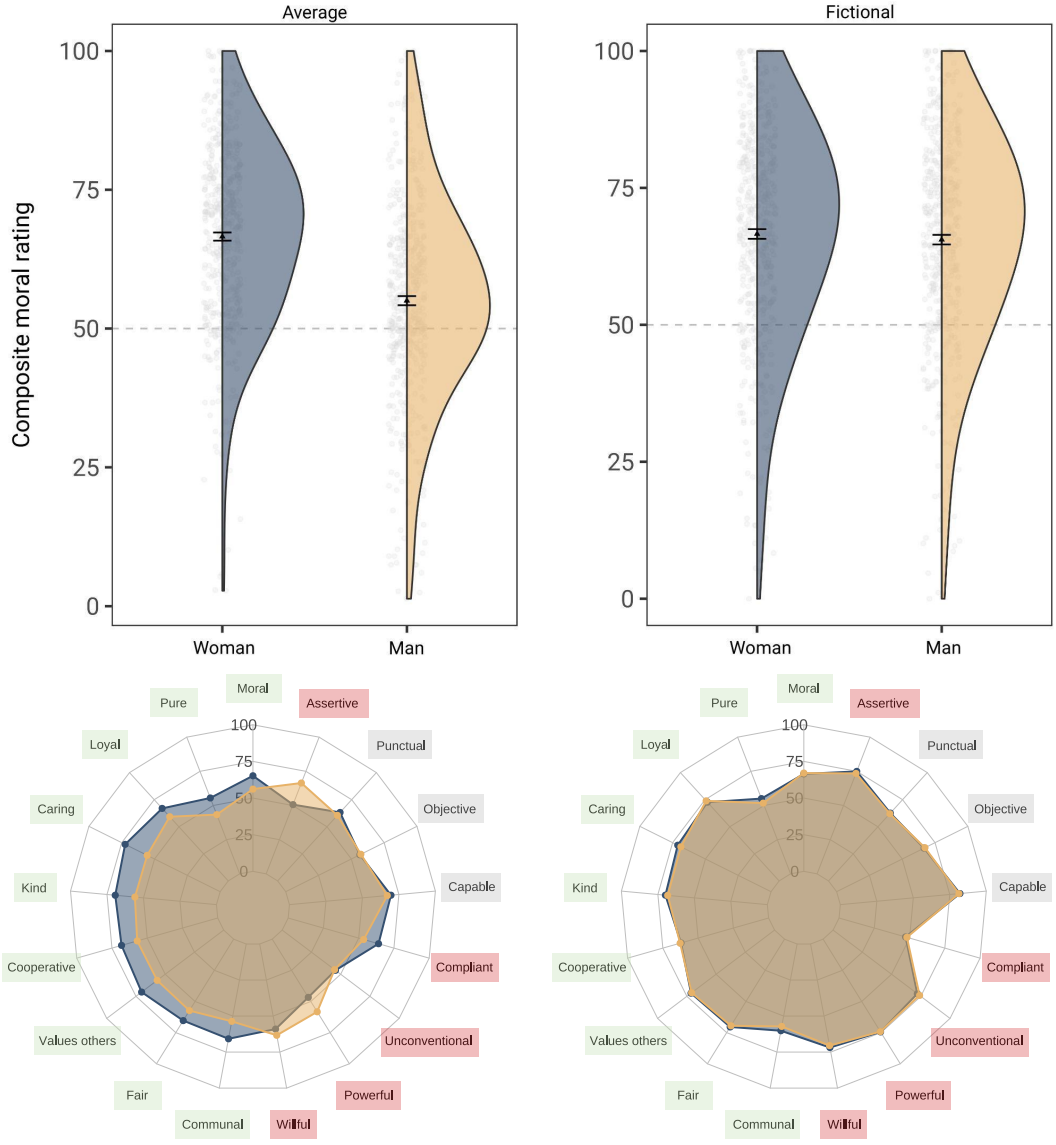


Figure 6. Top Panel: Composite moral ratings of each of the two gender conditions for average and fictional profiles. Average women were rated more morally good than average men, but there were no significant differences in fictional characters. Means are represented by black triangles. Error bars represent 95% Confidence Intervals. Bottom Panel: Radar plots depicting average individual trait ratings for Average (left side) and Fictional (right side) women and men profiles. Green labels represent moral traits, red labels represent agency traits of interest, and grey represent non-moral traits.

Looking at individual trait predictions, we largely found support for our predictions. The average woman profile was rated as more moral (using the overall morality item; b = 9.08, SE = 0.80, t = 11.35, *95% CI* [7.51, 10.65], $p < .001$, d = 0.51), more compliant (b = 10.58, SE = 0.95, t = 11.13, *95% CI* [8.72, 12.45], $p < .001$, d = 0.50), less powerful (b = -11.59, SE = 1.01, t = -11.51, *95% CI* [-13.56, -9.61], $p < .001$ d = -0.52), less willful (b = -4.33, SE = 0.91, t = -4.76, *95% CI* [-6.11, -2.54], $p < .001$ d = -0.21), and less assertive (b = -15.80, SE = 1.06, t = -14.88, *95% CI* [-17.89, -13.72], $p < .001$ d = -0.67) than the average man (see Supplemental Material for correlations). However, contrary to predictions, we did not find evidence of differences in unconventionality between average women and average men, p = .38 (for findings for each trait, see Supplement)[6].

Lastly, we preregistered examining whether there were differences in moral ratings of both fictional and average profiles based on *participant* gender. We predicted and found evidence of an interaction between participant gender and profile gender on ascriptions of moral traits using the composite measure (b = -3.31, SE = 1.32, t = -2.52, *95%CI* [-5.90, -0.73], p = .012, β = 0.17). However, the direction of this interaction was not as predicted: Women rate other women as more morally good than men do (see Supplemental Materials for full results).

Discussion

People ascribe more moral goodness—in the form of peacefulness, cooperation, forgiveness, valuing others, kindness, care and more—to women than to men. However, contrary to our predictions, these effects do not emerge for fictional characters, who are permitted more rule-breaking and badness than real people, both men and women.

Study 3

In Study 3, we move from traits to behaviors, testing whether moral goodness is associated with femininity in a large, publicly available set of 2,375 behavioral stimuli (the Delaware Behavior Database [DBD]; Mende-Siedlecki & Havlicek, 2025), which were previously normed along multiple dimensions relevant to social cognition and moral judgment. We tested whether actions rated as more morally good were also rated as more feminine. Secondarily, we examined a) whether a similar association exists with ratings of trustworthiness, b) whether these associations are conserved when controlling for potential confounds, and c) whether these associations are robust across a broad spectrum of behavior.

Method

Sample size and power analysis

[6] We also preregistered two additional tests. First, we preregistered and tested whether there was an interaction between average vs. fictional profiles and gender on compliance ratings. We do indeed find the predicted interaction. We report those results in full in the Supplemental Material. We also preregistered testing for similarity among the traits for full personality profiles, specifically whether the correlations among the different moral trait ratings were more similar to the correlations among the traits of the ideal person for average women compared to average men. Contrary to predictions, we did not find evidence of large differences when comparing the overall associations among traits for ideal, average, and different gender profiles. We report these data in full in the Supplemental Material.

A total of 2,876 participants were recruited from Amazon Mechanical Turk (MTurk) to complete the study. We conducted a sensitivity power analyses using G*Power (Faul et al., 2009). As in Study 2a, we specified a two-tailed test with alpha set to .05, and power set to 80%. This suggested that we were powered to detect a correlation of $r = 0.05$. The only inclusion criterion was current United States residence. Additional information about the sample is available in the Supplemental Material.

Stimuli

The behaviors in the set included extremely morally relevant behaviors like intentionally setting a house on fire, and behaviors that are not relevant to morality at all, like getting one's hair caught in a hair dryer. The behaviors also varied in valence, including relatively neutral, negative, and positive actions. Each participant was randomly assigned to characterize behaviors along one of these specific dimensions. Ratings were obtained across multiple rounds of data collection; in each case, participants were presented with a random subset of a larger set of 2375 behaviors to evaluate. On average, participants rated 246.65 behaviors, and each behavior obtained an average 29.89 ratings.

For the present analysis, we removed irrelevant stimuli from our analyses that represented experiences or things that happen to people (e.g., getting one's hair stuck in a hair dryer) rather than actions they do themselves. This left us with a total of 1,864 behaviors for further analysis. Additional information about the behaviors is available in the Supplemental Material.

Procedure & Design

Participants who helped characterize the DBD (N = 2,876 participants; 49% Female; 50% Male, 1% Other or missing; Mage = 36.63; SD_{age} = 11.66) each rated a randomly selected subset of 2,375 behaviors on gender stereotypicality (rated on a scale from 0 "very stereotypically male" to 8 "very stereotypically female"), moral relevance (0 "not at all morally relevant" to 8 "extremely morally relevant"), valence (-4 "very negative" to +4 "very positive"), trustworthiness rated on a scale from -4 "very untrustworthy" to +4 "very trustworthy"), and more. Participants also rated the behaviors on sociability, intelligence, capability, arousal, perceived frequency. We created a moral goodness score by multiplying ratings of valence by ratings of moral relevance, allowing us to identify the set of behaviors that were most moral and most positive. Other dimensions not relevant to the present analysis (e.g., dominance, status, race) were not analyzed and are not presented here.

Results

Our primary analyses were correlational. We first assessed the association between ratings of gender stereotypicality and the moral goodness composite, between gender stereotypicality and ratings of trustworthiness (typically considered a core moral trait; Brambilla et al., 2021) and sometimes measured as a proxy for impressions of morality (e.g., Mende-Siedlecki et al., 2013). Accordingly, we calculated Pearson correlation coefficients across the stimuli. Table 3 presents the correlation coefficients for each comparison.

As predicted, gender is positively correlated with moral goodness (i.e., with the valence by moral relevance term; $r = 0.31$, $95\%CI = [0.27, 0.35]$, $t(1862) = 14.07$, $p < .001$): the more morally good a behavior was rated, the more stereotypically feminine. A similar relationship emerged for trustworthiness ($r = 0.31$, $95\%CI = [0.27, 0.35]$, $t(1862) = 14.20$, $p < .001$), such that more feminine behaviors were also rated as being more trustworthy.

Table 3. Inter-correlations among predictors.

	Gender	Valence	Moral Relevance	Trustworthiness	Morality by Valence
1. Gender	—	.31***	-.06	.31***	.31***
2. Valence		—	-.08	.86***	.94***
3. Moral Relevance			—	-.15**	.03
4. Trustworthiness				—	.89***
5. Morality by Valence					—

Note. ***:** p < .05, **: p <.01, ***: p <.001 P-values are Holm adjusted.

Next, we used multiple regression to confirm whether the relationships examined in our primary correlational analyses were robust to potential confounds. For example, it might be the case that highly sociable behaviors tend to be seen as both more female-stereotypic and more morally good, relative to highly unsociable behaviors. To rule out these alternative explanations, we tested whether evaluations of gender stereotypicality were still associated with the moral goodness composite and trustworthiness even when controlling for arousal, perceived frequency, and evaluations of sociability, intelligence, and capability.

Accordingly, we pit gender prototypicality against a host of potential confounds in multiple regressions first predicting moral goodness and then, separately, trustworthiness ratings. In both cases, gender stereotypicality remained a significant predictor, even when controlling for other features that might have explained away these relationships. Specifically, moral goodness (b = 0.392, t = 5.270, p < .001; Table 4A) and trustworthiness (trustworthiness ratings (b = 0.050, t = 4.456, p < .001; Table 4B) were both positively associated with gender stereotypicality ratings when accounting for arousal, perceived frequency, sociability, intelligence, and capability as competing predictors.

Table 4. Multiple regression results pitting gender stereotypicality against possible confounds as predictors of trustworthiness and morality.

A. Gender stereotypicality remains positively associated with moral goodness

Predictor	*b*	*SE*	*t*	*95% CI*	*p*
Intercept	-5.844	0.490	-11.921	[-6.806, -4.883]	<.001
Gender stereotypicality	0.392	0.074	5.270	[0.246, 0.538]	<.001
Sociability	2.127	0.103	20.717	[1.925, 2.328]	<.001
Intelligence	1.797	0.166	10.841	[1.472, 2.122]	<.001
Capability	2.667	0.164	16.281	[2.346, 2.989]	<.001
Arousal	0.483	0.107	4.534	[0.274, 0.692]	<.001
Perceived frequency	0.003	0.006	0.585	[-0.008, 0.015]	.559

B. Gender stereotypicality remains positively associated with trustworthiness

Predictor	*b*	*SE*	*t*	*95% CI*	*p*
Intercept	-0.282	0.074	-3.789	[-0.428, -0.136]	<.001
Gender stereotypicality	0.050	0.011	4.456	[0.028, 0.072]	<.001
Sociability	0.227	0.016	14.592	[0.197, 0.258]	<.001
Intelligence	0.433	0.025	17.198	[0.383, 0.482	<.001
Capability	0.344	0.025	13.822	[0.295, 0.392]	<.001
Arousal	-0.065	0.016	-3.988	[-0.096, -0.033]	<.001
Perceived frequency	0.004	0.001	4.695	[0.002, 0.006]	<.001

Finally, since gender stereotypicality ratings were also correlated with evaluations of sociability, intelligence, and capability, we also tested whether gender stereotypicality was still associated with the moral goodness composite and trustworthiness regardless of the levels of these other traits. The relationship between moral goodness and gender stereotypicality significantly emerges for behaviors associated with high (p = .01) and low sociability (p < .001), high (p = .011) and low (p < .001) intelligence, and high (p = .001) and low (p < .001) capability. See Supplemental Materials for full analyses.

Discussion

Morally good actions tend to be seen as more feminine actions—even when controlling for other positive qualities and inferences central to person perception.

Study 4

Moral traits tend to be more diagnostic of character judgments than traits like warmth or competence (e.g., Brambilla et al., 2011; 2019; Brambilla & Leach, 2014; Luttrell et al., 2022). If women are seen as morally superior to men, then moral transgressions should be more diagnostic for women than for men, where diagnosticity reflects how much a behavior can reveal about a person's character given how descriptively normative it is (Mende-Siedlecki et a., 2013). Such a pattern would indicate both that immoral behaviors are more diagnostic for women, and that expectations of moral goodness may (ironically) constrain the behavior of women more than men (see also Prentice & Carranza, 2002; Biernat & Vescio, 2002; Glick & Fiske, 2001; Rudman et al., 2012a). Our primary hypothesis was that people would update their judgment more negatively for women than for men for the very same moral violations. Further, because prior research suggests that in-group members receive harsher punishment for moral transgressions than out-group members (Tang et al., 2023; Chawla et al., 2020), we also explored whether updating judgments would be harsher when made by one gender compared to another.

Method

Sample size and power analysis

We preregistered and recruited a sample of 1,604 participants (Mean age = 38.4, SD = 13.6, Gender: 49.6% women, 49.6% men, 0.81% other or missing) from Prolific. This sample size was based on an a *priori* power analysis conducted in G*Power (Faul et al., 2009). For 80% power on a two tail test with alpha set to .05 to detect an effect of d = 0.20 (based on pilot work), 394 participants per cell was required (788 total). We opted to round 788 up to 800 to account for any exclusions based on our preregistered criteria for a total of 800 participants. We recruit equal numbers of men and women for a total sample of 1600. No participants failed the attention check question.

Stimuli

We first selected all behaviors in the original set in Study 3, then we restricted to only those that were negative, rated as moderately morally relevant (from 6 to 7.1 range on a 8-point scale), and rated as neither stereotypically masculine or stereotypically feminine (from 3.25 to 4.75 on an 8-point scale). An example action is "Stole money from the tip jar at a coffee shop". The full list is available in the Supplemental Material.

Procedure & Design

We designed a one-shot updating paradigm to test whether the very same immoral action is more diagnostic for women than for men. Participants were randomly assigned to one of two gender conditions (man vs. woman) for the one-shot task. They first rated the morality of the average woman or man (depending on their condition) by answering "How morally good or bad do you

think the average [man / woman] is?" on a 101-point scale where 0 = *extremely morally bad* and 100 = *extremely morally good* and the anchor set to 50. Next, they rated the same target but were additionally given a single transgressive behavior to update based on. Participants saw one of fourteen possible behaviors all of which were selected from the normed DBD (Mende-Siedlecki & Havlicek, 2025) used in Study 3 (see Supplemental Materials for additional information). Participants answered: "How morally good or bad do you think this person is now?" using the same scale as above, but they were anchored based on where the individual participant put their average person rating (see Figure 7; see also Cusimano & Lombrozo, 2021). The experiment was programmed and displayed using *jsPsych* (De Leeuw, 2015).

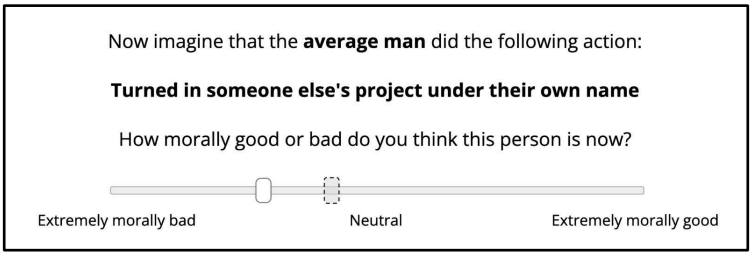


Figure 7. Example of an average man (vs. woman) condition updating trial of the Study 4 task. The updating judgment trial anchor started at the participant's baseline trial value and is depicted here in gray with a dashed border.

Results

We first tested whether baseline judgments differed between conditions. Replicating the general patterns from Studies 2a and 2b, we found that women were rated as more morally good than men, b = 8.49, SE = 0.85, t = 9.96, *95% CI* [6.82, 10.16], p < .001, d = 0.50.

Next, we conducted our preregistered analysis to test whether people update their estimation of women more than men following a transgression. To calculate the updating difference, we subtracted the baseline average judgment from the updated judgment. We specified a linear regression with the gender condition as the predictor. As predicted, the amount of change in judgments for women (M = -34.34, SD = 24.99) was more negative than those for men (M = -28.47, SD = 21.85), b = -5.93, SE = 1.17, t = -5.05, *95% CI* [-8.23, -3.62], p < .001, d = -0.25 (see Figure 8). Patterns for individual action stimuli are relatively consistent with the average and reported in the Supplemental Material. Additionally, because all of the updating actions were negative, we also took the absolute value to examine the magnitude change following the transgressive action. Results suggested that the magnitude of updating for women was greater than that for men, b = 6.30, SE = 1.10, t = 5.73, *95% CI* [4.14, 8.45], p < .001, d = 0.29. In other words, a single moral transgression is more costly for a woman's moral standing than a man's.

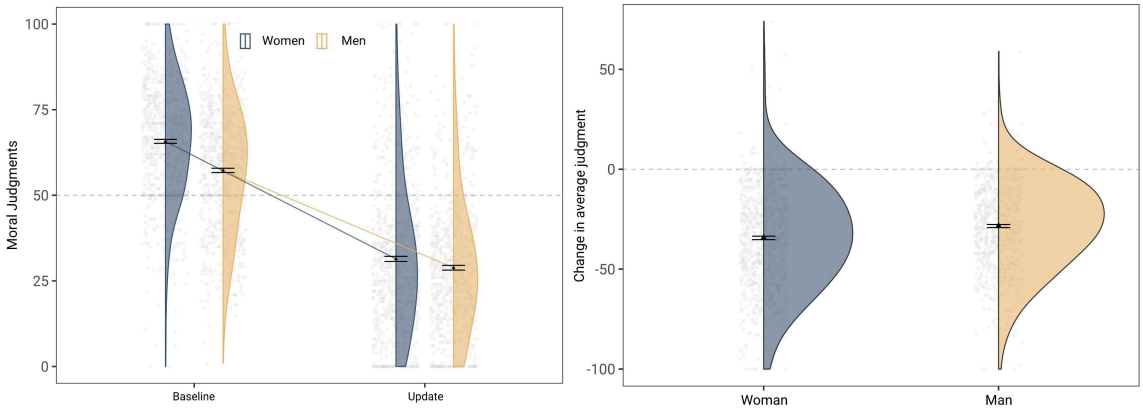


Figure 8. Updating judgments for each of the two gender conditions. Right Panel: Depicts average change from baseline to update judgment. Left Panel: Overall, updating judgments in the woman condition were more negative than in the man condition. Zero line indicates no update from a participant's baseline judgment. Means are represented by black triangles. Error bars represent 95% Confidence Intervals.

We also planned to explore whether updating differences vary by participant gender. As such, we specified a linear model that included participant gender, gender condition, and their interaction. There was again a main effect of the gender condition such that updates to women were more negative than those for men, b = -8.91, SE = 1.66, t = -5.35, *95% CI* [-12.17, -5.65], $p < .001$, β = 0.38. There was also a main effect of participant gender such that men updated less negatively than women, b = 3.95, SE = 1.65, t = 2.40, *95% CI* [0.72, 7.19], p = .017, β = 0.17. Critically, results also revealed a significant interaction, b = -6.07, SE = 2.35, t = -2.58, *95% CI* [-10.68, -1.45], p = .010, β = -0.26. Tukey's adjusted pairwise comparisons were used to decompose the interaction. Women in the woman-target condition updated more negatively than men in the woman-target condition (b = -8.91, SE = 1.66, $t(1587)$ = -5.35, $p < .001$). Similarly, women in the woman-target condition updated more negatively than men in the man-target condition (b = -6.79, SE = 1.67, $t(1587)$ = -4.07, p = .0003). Men in the woman-target condition updated more positively than women in the man-target condition (b = 4.96, SE = 1.66, $t(1587)$ = 2.99, p = .015). No other significant differences emerged. Women are both subject to harsher evaluations for moral transgressions and are more likely to dole them out.

Discussion

These findings suggest that impressions of women's moral character are more fragile than that of men's: a single transgression produces sharper declines in perceived moral goodness for women than men.

General Discussion

People think both that women are morally superior to men, and that—in contrast to work in moral and developmental psychology, which have historically ascribed greater moral reasoning ability to men than to women (Kohlberg 1976; but see Gilligan, 1982)—moral goodness is femininized. We found evidence for the gendered nature of moral goodness at the level of faces, traits, behaviors, and impressions: People represent the face of someone morally good as a woman's face. When people assign traits to others, people ascribe more morally good traits when they have a woman in mind both spontaneously and when prompted. Morally good behaviors are rated as more stereotypically feminine and finally, a single, identical moral transgression engenders more negative impression updating for women than for men.

These findings highlight that even moral goodness has its downsides (see Wolf, 1982). Indeed, Study 4 suggests that women face greater costs to their moral character than men when they break equivalent moral rules—or assert agency (Rudman et al., 2012b). The moral domain is full of rules, obligations, and principles (e.g., Cushman, 2013; Nichols & Mallon, 2006; Piaget, 1932/2013; Tomasello, 2020; Wylie et al., 2024)—which are usually beneficial (see Ellemers & van den Bos, 2012; Janoff-Bulman & Carnes, 2013; Baumard et al., 2013; Jordan, 2023; see also Wylie & Gantman, 2024)—but are at odds with autonomy and agency (Hahl et al., 2018; Keltner et al. 2003; Oostrom et al., 2021; van Kleef, 2023; Wylie et al., 2025). This tension, which bears more strongly on women than men, is relaxed in fiction (Study 2b). This is both an interesting avenue for future research, and reveals that our findings are not merely effects of positivity.

Previous research on androcentrism has suggested that people tend to associate genderless generics like “human” with men (Bailey & LaFrance, 2017; Bailey et al., 2020; Bailey et al., 2022; see also Bailey et al., 2019). And humanness judgments are linked with morality (see Bastian et al. 2011; Haslam, 2006; Halsman & Bain, 2007) and gender (Martin & Mason, 2022). Our findings that women are more related to moral goodness than men sits uneasily with these findings, and may pose an interesting paradox.

This work does not directly speak to *where* these perceptions come from. Perhaps these perceptions simply come from everyday experience (e.g., descriptive norms). Women are often nice to people (Eagly & Hall, 2025)., and people remember them that way (Markowitz et al., 2025). Yet, Study 4 suggests a prescriptive component—if statistical norms alone drove updating, we would expect to see similar updating for men and women, given that the transgressions were identical.

One limitation of the present research is the exclusive US context. Future research should investigate how cultural and situational factors—such as tightness (Gelfand et al., 2011; see Qin et al., 2023) and moments of gender backlash (Faludi, 1991; see also Rudman et al., 2012a)—influence the link between moral goodness and femininity. Additionally, research suggests that ideas of masculinity often explicitly forbid feminine behavior (Dahl et al., 2015). Future research should explore whether men avoid or are barred from moral goodness because it is feminized.

People think women are morally superior to men. We find an association between moral goodness and femininity at the level of mental representations of faces, personality traits, individual behaviors, and overall character judgments, and is present when both men and women make judgments.

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. Journal of personality and social psychology, 93(5), 751.

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. In Advances in experimental social psychology (Vol. 50, pp. 195-255). Academic Press.

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological review*, *128*(2), 290.

Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. Trends in cognitive sciences, 21(2), 69-79.

Atari, M., Mehl, M. R., Graham, J., Doris, J. M., Schwarz, N., Davani, A. M., ... & Dehghani, M. (2023). The paucity of morality in everyday talk. Scientific Reports, 13(1), 5967.

Badgett, M. V. & Folbre, N. (1999). Assigning care: Gender norms and economic outcomes. Int'l Lab. Rev., 138, 311.

Bailey, A. H., & LaFrance, M. (2017). Who counts as human? Antecedents to androcentric behavior. Sex Roles, 76, 682-693.

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2019). Is man the measure of all things? A social cognitive account of androcentrism. Personality and social psychology review, 23(4), 307-331.

Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2020). Implicit androcentrism: Men are human, women are gendered. Journal of Experimental Social Psychology, 89, 103980.

Bailey, A. H., Williams, A., & Cimpian, A. (2022). Based on billions of words on the internet, people= men. Science Advances, 8(13), eabm2463.

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. Behavioral and Brain Sciences, 36(1), 59-78.

Bareket, O., & Fiske, S. T. (2023). A systematic review of the ambivalent sexism literature: Hostile sexism protects men's power; benevolent sexism guards traditional gender roles. Psychological bulletin.

Bastian, B., Laham, S. M., Wilson, S., Haslam, N., & Koval, P. (2011). Blaming, praising, and protecting our humanity: The implications of everyday dehumanization for judgments of moral status. British Journal of Social Psychology, 50(3), 469-483.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. convergence, 12(1), 2.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition*, 167, 25-37.

Becker, J. C., & Wright, S. C. (2011). Yet another dark side of chivalry: Benevolent sexism undermines and hostile sexism motivates collective action for social change. Journal of personality and social psychology, 101(1), 62.

Biernat, M., & Vescio, T. K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. Personality and Social Psychology Bulletin, 28(1), 66-77.

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. Social Cognition, 32(4), 397-408.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. European journal of social psychology, 41(2), 135-143.

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. Journal of Experimental Social Psychology, 82, 64-73.

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. In Advances in experimental social psychology (Vol. 64, pp. 187-262). Academic Press.

Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. Psychological science, 28(1), 92-103.

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. Personality and social psychology bulletin, 25(4), 413-423.

Charlesworth, T. E., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. Psychological Science, 32(2), 218-240.

Chawla, M., Earp, B. D., & Crockett, M. J. (2020). A neuroeconomic framework for investigating gender disparities in moralistic punishment. Current Opinion in Behavioral Sciences, 34, 166-172.

Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. The evolution of morality, 27-51.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. Personality and social psychology review, 17(3), 273-292.

Cusimano, C., & Lombrozo, T. (2021). Morality justifies motivated reasoning in the folk ethics of belief. Cognition, 209, 104513.

Daalmans, S., & Hendrix, J. (2024). Gender, Identity, and Entertainment. Entertainment Media and Communication, 20, 429.

Daalmans, S., Kleemans, M., & Sadza, A. (2017). Gender representation on gender-targeted television channels: A comparison of female-and male-targeted TV channels in the Netherlands. Sex roles, 77, 366-378.

Dahl, J., Vescio, T., & Weaver, K. (2015). How threats to masculinity sequentially cause public discomfort, anger, and ideological dominance over women. Social Psychology.

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: consequences for women's performance. Journal of personality and social psychology, 93(5), 764.

Deci, E. L., & Ryan, R. M. (2013). The importance of autonomy for development and well-being. Self-regulation and autonomy: Social and developmental dimensions of human conduct, 19-46.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behavior research methods, 47, 1-12.

Diekman A. B., Eagly A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. Personality and Social Psychology Bulletin, 26, 1171–1188.

Dotsch, R. (2023). rcicr: Reverse-correlation image-classification toolbox [R package Version 1.0.1]. Retrieved from https://CRAN.R-project.org/package=rcicr

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. Social Psychological & Personality Science, 3, 562–571.

Dotsch, R., Wigboldus, D. H. J., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. Psychological Science, 19, 978–980.

Eagly, A.H. (1987). Sex Differences in Social Behavior: A Social-role interpretation (1st ed.). Psychology Press. https://doi.org/10.4324/9780203781906

Eagly, A. H. (2009). The his and hers of prosocial behavior: an examination of the social psychology of gender. American psychologist, 64(8), 644.

Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. Journal of personality and social psychology, 60(5), 685.

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. European review of social psychology, 5(1), 1-35.
Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. Journal of personality and social psychology, 46(4), 735.
Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? An analysis of attitudes, beliefs, and emotions. Psychology of women quarterly, 15(2), 203-216.
Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. American psychologist, 75(3), 301.
Eagly, A. H., & Hall, J. A. (2025). The kernel of truth in gender stereotypes: Consider the avocado, not the apple. Journal of Experimental Social Psychology, 118, 104713.
Ellemers, N. (2018). Gender stereotypes. Annual review of psychology, 69(1), 275-298.
Ellemers, N., & van den Bos, K. (2012). Morality in groups: On the social-regulatory functions of right and wrong. Social and Personality Psychology Compass, 6(12), 878-889.
Faludi, S. (1991). Backlash: The Underclared War Against Women. Vintage Books.
Faul, F., Erdfelder, E., Buchner, A. et al. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods 41, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
Fischer, R., & Boer, D. (2011). What is more important for national well-being: money or autonomy? A meta-analysis of well-being, burnout, and anxiety across 63 societies. Journal of personality and social psychology, 101(1), 164.
Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current directions in psychological science*, *27*(2), 67-73.
Flores-Robles, G., & Gantman, A. P. (2024). Notions of Care Labor Are Antithetical to Profitable Labor. Psychology of Women Quarterly, 48(4), 475-490.
Foster-Hanson, E., & Rhodes, M. (2023). Stereotypes as prototypes in children's gender concepts. Developmental science, 26(4), e13345.
Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. Journal of personality and social psychology, 101(1), 109.
Gelfand, M., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'amato, A., Ferrer, M. S., Fischlmayr, R. C.…Yamaguchi, S. (2011). Differences Between Tight and Loose Cultures: A 33-Nation Study. Science, 322, 1100-1104.
Gilligan, C. (1982). In a different voice: Psychological theory and women's development. Harvard University Press
Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. American psychologist, 56(2), 109.
Glick, P., & Fiske, S. T. (2018). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. In Social cognition (pp. 116-160). Routledge.
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.
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.
Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. Journal of personality and social psychology, 101(2), 366.
Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In Advances in experimental social psychology (Vol. 47, pp. 55-130). Academic Press.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. Psychological inquiry, 23(2), 101-124.2

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. Journal of Experimental Psychology: General, 143(4), 1600.

Hahl, O., Kim, M., & Zuckerman Sivan, E. W. (2018). The authentic appeal of the lying demagogue: Proclaiming the deeper truth about political illegitimacy. American Sociological Review, 83(1), 1-33.

Hamilton, M. C., Anderson, D., Broaddus, M., & Young, K. (2006). Gender stereotyping and under-representation of female characters in 200 popular children's picture Books: A twenty-first century update. Sex Roles, 55(11–12), 757–765. https://doi.org/10.1007/s11199-006-9128-6

Haslam, N. (2006). (PDF) Dehumanization: An Integrative Review. ResearchGate. https://www.researchgate.net/publication/6927454_Dehumanization_An_Integrative_Review

Haslam, N., & Bain, P. (2007). Humanizing the self: Moderators of the attribution of lesser humanness to others. Personality and Social Psychology Bulletin, 33(1), 57-68.

Heilman, M. E. (1983). Sex bias in work settings: The Lack of Fit model. *Research in Organizational Behavior, 5,* 269–298.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. Journal of social issues, 57(4), 657-674.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. Research in organizational Behavior, 32, 113-135.

Heilman, M. E., Caleo, S., & Manzi, F. (2024). Women at work: pathways from gender stereotypes to gender bias and discrimination. Annual Review of Organizational Psychology and Organizational Behavior, 11(1), 165-192.

Hideg, I., & Ferris, D. L. (2016). The compassionate sexist? How benevolent sexism promotes and undermines gender equality in the workplace. Journal of personality and social psychology, 111(5), 706.

Hsu, N., Badura, K. L., Newman, D. A., & Speach, M. E. P. (2021). Gender,"masculinity," and "femininity": A meta-analytic review of gender differences in agency and communion. Psychological Bulletin, 147(10), 987.

Jackson, J. C., Halberstadt, J., Takezawa, M., Liew, K., Smith, K., Apicella, C., & Gray, K. (2023). Generalized morality culturally evolves as an adaptive heuristic in large social networks. Journal of Personality and Social Psychology.

Janoff-Bulman, R., & Carnes, N. C. (2013). Surveying the moral landscape: Moral motives and group-based moralities. Personality and Social Psychology Review, 17(3), 219-236.

Jordan, J. J. (2023). A pull versus push framework for reputation. Trends in Cognitive Sciences, 27(9), 852–866. https://doi.org/10.1016/j.tics.2023.06.005

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. Psychological review, 110(2), 265.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-development approach. Moral development and behavior: Theory research and social issues, 31-53.

Krys, K., Capaldi, C. A., Van Tilburg, W., Lipp, O. V., Bond, M. H., Vauclair, C. M., ... & Ahmed, R. A. (2018). Catching up with wonderful women: The women-are-wonderful effect is smaller in more gender egalitarian societies. International Journal of Psychology, 53, 21-26.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. Journal of statistical software, 82(13).

LaFrance, M., & Hecht, M. A. (1995). Why smiles generate leniency. Personality and Social Psychology Bulletin, 21(3), 207-2.

LaFrance, M., & Hecht, M. A. (1995). Why smiles generate leniency. Personality and Social Psychology Bulletin, 21(3), 207-214.
LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. Psychological bulletin, 129(2), 305.
Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. Advances in methods and practices in psychological science, 1(2), 259-269.
Landy, J. F., Piazza, J., & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. Personality and social psychology bulletin, 42(9), 1272-1290.
Long JA (2022). jtools: Analysis and Presentation of Social Scientific Data_. R package version 2.2.0, https://cran.r-project.org/package=jtools .
Luttrell, A., Sacchi, S., & Brambilla, M. (2022). Changing impressions in competence-oriented domains: The primacy of morality endures. Journal of Experimental Social Psychology, 98, 104246.
Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. Behavior research methods, 47, 1122-1135.
Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: agentic and communal differences. Journal of Applied Psychology, 94(6), 1591.
Manne, K. (2017). Down Girl: The Logic of Misogyny. In Oxford Scholarship Online. Oxford University Press. https://doi.org/10.1093/oso/9780190604981.001.0001
Martin, A. E., & Slepian, M. L. (2021). The primacy of gender: Gendered cognition underlies the big two dimensions of social cognition. Perspectives on Psychological Science, 16(6), 1143-1158.
Martin, A. E., & Mason, M. F. (2022). What does it mean to be (seen as) human? The importance of gender in humanization. Journal of Personality and Social Psychology, 123(2), 292.
Markowitz, D. M., Mazzuchi, T., Syropoulos, S., Law, K. F., & Young, L. (2025). An exploration of basic human values in 38 million obituaries over 30 years. *Proceedings of the National Academy of Sciences*, *122*(35), e2510318122.
Martin, A. E., & Mason, M. F. (2022). What does it mean to be (seen as) human? The importance of gender in humanization. *Journal of Personality and Social Psychology*, *123*, 292-315.
Melnikoff, D. E., & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. Proceedings of the National Academy of Sciences, 115(4), E592-E600.
Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. Journal of Neuroscience, 33(50), 19406-19415.
Mende-Siedlecki, P., & Havlicek, L. (2025). The Delaware Behavior Database: A set of social behaviors and corresponding norming data. https://doi.org/10.31234/osf.io/b37rz
Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: status incongruity and backlash against modest men. Psychology of Men & Masculinity, 11(2), 140.
Niemi, L., & Young, L. (2013). Caring across boundaries versus keeping boundaries intact: Links between moral values and interpersonal orientations. PloS one, 8(12), e81605.
Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. Cognition, 100(3), 530-542.
Oostrom, J. K., Ronay, R., & Van Kleef, G. A. (2021). The signalling effects of nonconforming dress style in personnel selection contexts: do applicants' qualifications matter?. European Journal of Work and Organizational Psychology, 30(1), 70-82.
Packer, D. J. (2008). On being both with us and against us: A normative conflict model of dissent in social groups. Personality and social psychology review, 12(1), 50-72.
Piaget, J. (1932/2013). The Moral Judgment of the Child. https://doi.org/10.4324/9781315009681

Pietaryte, K., & Suzina, A. C. (2023). Female representation in Netflix Global Original programming: A comparative analysis of 2019 drama series. Critical Studies in Television, 18(1), 41-61.
Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. Psychology of women quarterly, 26(4), 269-281.
Qin, X., Chua, R. Y., Tan, L., Li, W., & Chen, C. (2023). Gender bias in cultural tightness across the 50 US states, its correlates, and links to gender inequality in leadership and innovation. PNAS nexus, 2(8), pgad238.
R Core Team, R. (2013). R: A language and environment for statistical computing.
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.
Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. Personality and social psychology bulletin, 26(4), 419-435.
Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. Journal of Social Issues, 57(4), 743–762. https://doi.org/10.1111/0022-4537.00239
Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men?. Journal of personality and social psychology, 87(4), 494.
Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. Personality and Social Psychology Bulletin, 26(11), 1315–1328. https://doi.org/10.1177/0146167200263001
Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. Research in organizational behavior, 28, 61-79.
Rudman, L. A., Moss-Racusin, C. A., Glick, P., & Phelan, J. E. (2012a). Reactions to vanguards: Advances in backlash theory. In Advances in experimental social psychology (Vol. 45, pp. 167-227). Academic Press.
Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012b). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of experimental social psychology*, *48*(1), 165-179.
Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. Personality and Social Psychology Review, 22(1), 32-70.
Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?. Journal of Research in Personality, 47(5), 609-612.
Sczesny, S., Nater, C., & Eagly, A. H. (2018). Agency and communion: Their implications for gender stereotypes and gender identities. In Agency and communion in social psychology (pp. 103-116). Routledge.
Stuhler, O. (2024). The gender agency gap in fiction writing (1850 to 2010). *Proceedings of the National Academy of Sciences*, *121*(29), e2319514121.
Sun, J., & Goodwin, G. P. (2020). Do people want to be more moral?. Psychological science, 31(3), 243-257.
Sun, J., Wu, W., & Goodwin, G. (2022, July 13). Are Moral People Happier? Answers from Reputation-Based Measures of Moral Character. https://doi.org/10.31234/osf.io/sd8t4
Tang, S., Shepherd, S., & Kay, A. C. (2023). Morality's role in the Black Sheep Effect: When and why ingroup members are judged more harshly than outgroup members for the same transgression. European Journal of Social Psychology, 53(7), 1605-1622.
Tomasello M. (2020) The moral psychology of obligation. Behavioral and Brain Sciences 43, e56: 1–58. doi:10.1017/S0140525X19001742.

Tuchman, G. (2000). The symbolic annihilation of women by the mass media. In Culture and politics: A reader (pp. 150-174). New York: Palgrave Macmillan US.
Vaage, M. B. (2015). The antihero in American television (1st ed.). Routledge. https://doi.org/10.4324/9781315715162
Van de Vondervoort, J. W., & Hamlin, J. K. (2017). Preschoolers' social and moral judgments of third-party helpers and hinderers align with infants' social evaluations. Journal of Experimental Child Psychology, 164, 136–151. https://doi.org/10.1016/j.jecp.2017.07.004
van Kleef, G. A. (2023). When and how norm violators gain influence: Dominance, prestige, and the social dynamics of (counter) normative behavior. Social and Personality Psychology Compass, 17(5), e12745.
Vescio, T. K., Schermerhorn, N. E., Lewis, K. A., Yamaguchi-Pedroza, K., & Loviscky, A. J. (2023). Masculinity threats sequentially arouse public discomfort, anger, and positive attitudes toward sexual violence. Personality and social psychology bulletin, 01461672231179431.
Vescio, T. K., Schlenker, K. A., & Lenes, J. G. (2010). Power and sexism. In A. Guinote & T. K. Vescio (Eds.), The social psychology of power (pp. 363–380). The Guilford Press.
Williams, J., & Best, D. (1990). Measuring sex stereotypes: A multinational study. Beverly Hills, CA: Sage.
Wolf, S. (1982). Moral saints. The Journal of Philosophy, 79(8), 419-439.
Wylie, J., & Gantman, A. (2024). Cooperation, domination: Twin functions of third-party punishment. Social and Personality Psychology Compass, 18(8), e12992.
Wylie, J., Milless, K. L., Sciarappo, J., & Gantman, A. (2024). The biased enforcement of rarely followed rules. Personality and Social Psychology Bulletin, 01461672241252853.
Wylie, J., Lindauer, M., & Gantman, A. (2025). People can find their true selves outside moral pursuits. *Cognition*.