The Influence of Gender on Accuracy in Estimating Moral Behavior

Tori N. LeVier[1], David Tannenbaum[2], Ana Gantman[3,4], Paul Bloom[5] & Jordan Wylie (jordan.wylie@cornell.edu)[1]

[1]Department of Psychology, Cornell University
[2]Department of Management, University of Utah
[3]Department of Psychology, Brooklyn College
[4]Department of Psychology, CUNY Graduate Center
[5]Department of Psychology, University of Toronto

Abstract

Accurate beliefs about the social world are essential to adaptive social functioning. However, people's perceptions of the social world are often inaccurate. Across three experiments (N = 991), we examine people's estimates of the prevalence of ethical and unethical behaviors for men and women. In Study 1, participants reported their own engagement in common unethical behaviors (e.g., cheating on a romantic partner), and then estimated how often men and women engaged in those same behaviors. As predicted, people overestimated the frequency of unethical behavior—an effect which was amplified for men. In Study 2, participants again overestimated the frequency of unethical behaviors, but also underestimated the frequency of ethical behaviors. These effects were again amplified for estimates of men's behaviors. Study 3 replicated these findings using hypothetical scenarios. Overall, we find consistent evidence of moral cynicism concerning the behaviors of others. People systematically assume more wrongdoing and less virtue than actually exists.

Keywords: estimation; moral behavior; social judgment

Introduction

The capacity to form beliefs about the structure of the world around us is critical for navigating everyday life. But these beliefs are often inaccurate. For instance, people systematically overestimate the size of small demographic groups and underestimate the size of large groups, regardless of the content of those estimates (e.g., racial, nonracial, nonpolitical, out-groups, etc.; Guay et al., 2025). That is, people are prone to making systematic estimation errors, which may reflect greater limitations related not just to probability, but also possibility (Shtulman & Phillips, 2018).

The consistency and breadth of these errors is puzzling. Accurate beliefs about the world, especially the social world, confer many advantages. They allow people to make effective decisions, coordinate with others, avoid conforming to bad norms (Prentice & Miller, 1993, 1996), and to accurately anticipate risks and rewards (e.g., Johnson-Laird, 2004; Bosworth, 2017; Tetlock, 2017). Why then do so many fail to form accurate beliefs about the social world around them?

One reason relates to social categories. People often form expectations about the world around them based on social category information (e.g., Rhodes & Baron, 2019), and those expectations can be biased or inaccurate. Indeed, past research suggests that inaccurate beliefs about others have consequences: For instance, when people overestimate the prevalence of a behavior like binge drinking (Haines & Spear, 1996) or a driving violation (e.g., speeding; Manstead et al., 1992), they are more likely to engage in that behavior themselves (e.g., Prentice & Miller, 1993; Haines & Spear, 1996). Holding erroneous estimates about the prevalence of bad behaviors can license people to engage in them (e.g., Cialdini et al., 1991; Schultz et al., 2007), and social category information may be one especially salient cue that guides expectations.

Gender is a particularly salient social category as it tends to be visible. And there is a large body of literature which suggests that people have strong expectations about others based on their gender: Decades of scholarship suggest that people see women more positively (Eagly & Steffen, 1984), warm and communal (e.g., Eagly & Karau, 1991; Flores-Robles & Gantman, 2024; Glick & Fiske, 2001; see Hsu et al., 2021 for review), and morally good (Rudman & Glick, 2001; Torgler & Valev, 2006; Wylie et al., 2025) than they see men. These gender stereotypes are essentialized and early emerging (Meyer & Gelman, 2016; Foster-Hanson & Rhodes, 2023). In fact, men are often treated as a cognitive default or reference category in social judgment (Lei et al., 2022; Bailey et al., 2019). Gender provides an entrenched and readily available inference cue for a range of social judgments and predictions.

Here, we integrate across two rarely overlapping literatures to ask whether people's gendered expectations about moral behavior are accurate. We focus on morally relevant behavior because such behavior is especially consequential for social life. Morality is central to how people view and understand the self (see Strohminger, 2018), and how people judge others (e.g., Goodwin et al., 2014; Landy et al., 2016; Brambilla, 2021). Predictions about others' moral behavior also have important implications—research suggests that perceptions of how common a given prosocial behavior is affects the likelihood someone engages in it (Croson et al., 2009).

Across three experiments we examine the accuracy of predictions of moral behavior across gender. Since men and women each make up roughly half of the population, systematic over and underestimation biases associated with large and small groups (Guay et al., 2025) should not impact overall predictions for either gender. That is, gender is one social category where cognitive biases likely play a smaller role than estimates of other social categories (e.g., race).

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Overall, we hypothesized that there would be broad errors in the perceived frequency of the moral behavior of others, especially men. We expected that errors, both in underestimating good and overestimating bad behavior, would be greater for men because research suggests that people tend to see women as more moral than men (Wylie et al., 2025), and other work suggests that men make up the majority of negative news and even make up the majority of villains in fiction and entertainment (e.g., Daalmans et al., 2017). In Experiment 1, participants were asked to estimate how often men and women engaged in a set of unethical behaviors, as well as self-reporting how often they had. In Experiment 2, we replicate and extend the findings to include ethical behaviors as well. In Experiment 3, we measured participant predictions and responses to hypothetical unethical behaviors. All experiments were approved by our Institutional Review Board and Study 2 and Study 3 were pre-registered, and all code, materials, and data are available on the project's OSF page[1] (LeVier & Wylie, 2026).

Study 1

Study 1 tested whether participants' estimates of the frequency of engaging in unethical behavior were conditional on the target's gender. We then compared those estimates to the true rates of self-reported unethical behavior. We predicted that people would overestimate the rate of unethical behavior, and that this effect would be most pronounced when estimating the behavior of men.

Method

We recruited a sample of 213 (61% men, 39% women, M_{age} = 23 years old, *range*: 19–42 years) students from University of Utah. We ran the study over two academic semesters. Participants completed the study in exchange for course credit.

We provided men and women with a series of relatively common unethical behaviors ranging in severity (e.g., "have you ever copied the answers from another student during a test?", and "have you ever cheated on a romantic partner by having sex with another person?"). There were 25 behaviors in total that the research team created. Overall, the most common unethical behavior that people reported engaging in was lying to their parents (97.4%), and the least common was intentionally lying on an insurance claim (6.1%). Participants completed two tasks using these stimuli. First, participants indicated whether they have ever engaged in each unethical behavior (self-report task). Behaviors were presented in random order, one at a time, and were answered with yes/no responses. We refer to the results of this task as *empirical estimates*. Second, participants estimated the percentage (from 0 – 100%) of men and women in the sample (prediction task) who have engaged in each behavior, presented in a randomized order. We refer to the results of this task as the *predicted estimates*. The order of the two tasks was counterbalanced across participants.

For the self-report task, we told participants that they might encounter sensitive topics but that their responses would remain anonymous and de-identified. As an incentive for accurate responding, we notified participants at the beginning of the prediction block that the most accurate participant in the study would receive a $50 bonus payment. Participants were told that accuracy was measured as the mean squared difference between their estimates and actual response rates (Brier, 1950).

Results

We assessed judgment accuracy by comparing predicted responses to actual responses. For all analyses below, we coded self-report data as 1 when responding "Yes" and 0 when responding "No" to engaging in an unethical behavior, and prediction data was coded as falling between [0,1]. When the dependent variable was Yes/No self-reports, we used logistic regression; when the dependent variable was prediction data or calibration scores, we used linear regression. For all analyses we included gender of the target as a fixed effect. For the majority of analyses, we averaged the 25 unethical behaviors into a single estimate/prediction. However, when we treat each behavioral estimate separately, we get the same general pattern of results.

Real and Predicted Gender Differences We first assessed participants' predictions of men's and women's engagement in unethical behaviors using a mixed-effects model with target gender as a fixed effect and by-participant and by-behavior random intercepts included. Results suggested that people expect men to engage in unethical behavior more frequently than women, $\beta = 0.20$, SE = 0.01, 95% CI [0.17, 0.22], $t = 14.17$, $p < .001$. We also found that participant gender did not influence prediction estimates, $p = .546$.

Participants predicted that men, on average, responded "Yes" to engaging in unethical behavior 50.5% of the time (predicted estimate), when in reality they only reported engaging in unethical behaviors 42.0% of the time (empirical estimate) (see Figure 1a). Additionally, women were expected to respond "Yes" 45.1% of the time (predicted estimate), but only did so 41.3% of the time (empirical estimate) (see Figure 1a). Overall, participants *expected* men to be unethical more frequently than they expected women (5.4 percentage points; $t(24) = 7.25$, $p < .001$).

The actual difference between men and women was on average less than a single percentage point, and consistent with a null hypothesis of no gender difference ($p = 0.640$). Further, when looking at the difference between predicted and empirical estimates, we found that people systematically overestimated gender differences in unethical behavior ($t(24) = 2.89$, $p < .001$). On average, participants overestimated the difference between men and women by a factor of eight. We also examined whether participants were miscalibrated about men, women, or both. To do this, we compared predictions of men's responses to actual men's

[1] https://osf.io/hpwna/overview?view_only=6e882d8121ffa32af61bc3567170868

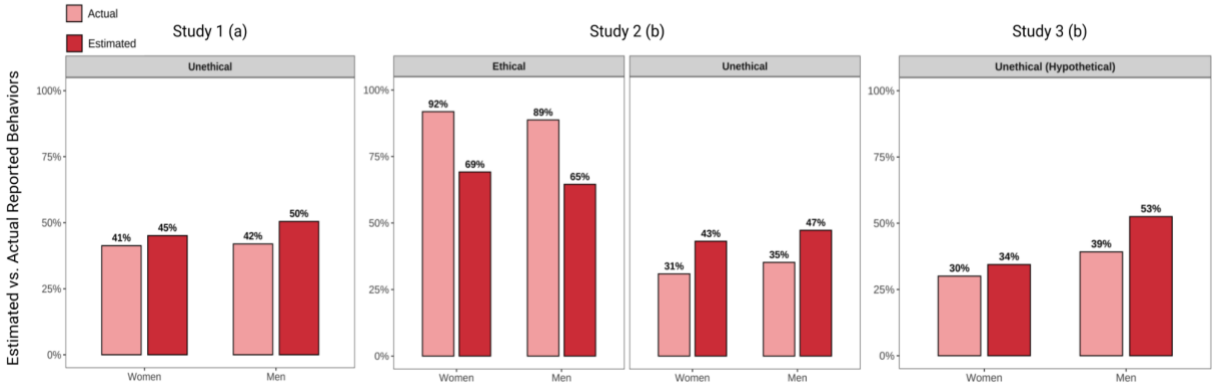

Figure 1: a) Rates of predicted behaviors vs. actual reported unethical behaviors in Study 1; b) Rates of predicted behaviors vs. actual ethical and unethical behaviors in Study 2; c) Rates of predicted vs. reported in hypothetical unethical scenarios

responses, and predictions of women's responses to actual women's responses. On average, participants overestimated men's reporting of unethical behavior by 8.5 percentage points ($t(212) = 8.52$, $p < .001$) and overestimated women's reporting of unethical behavior by 3.8 percentage points ($t(212) = 4.08$, $p < .001$). These results are consistent with prior research on naive cynicism (e.g., Kruger & Gilovich, 1999; Epley et al., 2006). People tended to overestimate the unethical behavior of others, and this bias was more pronounced when estimating men's behavior than women's behavior.

Judgment Discrimination Another form of accuracy is judgment discrimination (Murphy, 1973). Discrimination provides a measure of a forecaster's ability to differentiate likely from unlikely outcomes across trials. For instance, a participant may systematically overestimate unethical behavior (poor calibration) but could still reliably differentiate instances where response rates will be high rather than low (high discrimination). In this way, discrimination provides a measure of the information contained in a given judgment.

We test for discrimination by examining the correlation between predicted and actual reporting rates across participants and across questions (a similar pattern emerges when looking across participants but within questions, or when looking within participants but across questions). Predictions of men's response rates were correlated with how men actually responded ($r = 0.44$, 95% CI [0.42, 0.46]), and predictions of women's response rates were correlated with how women actually responded ($r = 0.49$, 95% CI [0.47, 0.51]). However, predictions of gender differences were not reliably correlated with actual gender differences ($r = -0.01$, 95% CI [−0.03, 0.02]). So although participants were moderately discriminating in their predictions of male and female behavior from question to question, they did not meaningfully differentiate between instances when the actual gender gap in unethical behavior was large or small.

Interim Discussion

Overall, participants predicted that both men and women engaged in more unethical behaviors than they actually did. This discrepancy between perceived and actual unethical behavior was overestimated to an even greater extent for men. However, the participants from Study 1 were a college sample, estimating rates of unethical behavior on their college campus. It is possible that this population has fewer opportunities for immoral behavior than outside of college campuses. Additionally, Study 1 only included unethical behaviors. However, accurate estimates of both ethical and unethical behavior are critical for adaptive social functioning. There is also a large body of literature which suggests that men are seen as more agentic than women (e.g., Stuhler, 2024; Hsu et al., 2021 for review), and overrepresented as extremely morally good (e.g., Daalmans et al., 2017). If these perceptions of agency are the driving factor, men should be expected to engage in more behaviors overall, whether they are ethical or unethical. However, if the driving factor is more related to expectations of women, then people should expect them to engage in fewer unethical behaviors and more ethical behaviors than men. In Study 2, we sought to replicate and extend the results from Study 1 to a representative sample and to include estimates of ethical behaviors too.

Study 2

Study 2 examined the accuracy of perceived gender differences for both unethical and ethical behaviors. We predicted that participants would continue to overestimate rates of unethical behavior on average, and especially for men. We also predicted that people would underestimate rates of ethical behavior, again especially for men. We predicted there would be no differences in actual behavior.

Method

We preregistered recruiting a US nationally representative sample of 496 participants from Prolific (49.8% men, 50.2% women, M_{age} = 46 years old, *range*: 20-86 years) with our sample size based on an a priori power analysis using G*power (Faul et al., 2007) and removed 4 for failing attention checks and missing self-reported gender values.

Participants were randomly presented with either 25 ethical or 25 unethical behaviors from a normed set of 150 behaviors. These behaviors were selected from a set of behaviors created by the research team to reduce confounds that might have been present in Study 1, and then normed with an independent sample. We selected stimuli for inclusion that were rated as morally relevant and gender-neutral and selected actions to ensure authority, care, fairness, honesty, and loyalty were represented in the set (see Graham et al., 2011). Additionally, we selected behaviors that have harm for our unethical behaviors (see e.g., Gray et al., 2014). Overall,

the most common unethical behavior that people reported engaging in was not following parents' curfew (66.5%), the least common was keeping a lost wallet (9.7%). The most common ethical behavior was doing one's part in a group project (98.5%), the least common was finding the owner of a lost wallet (70.0%).

We used the same procedure as in Study 1. Participants were again told at the beginning of the study that the most accurate responses would be placed in a lottery for a $5 bonus payment.

Results

As in Study 1, we first specified a linear mixed-effects model with target gender and moral valence as fixed-effects, and by-participant and by-behavior random intercepts. Results suggested that people expected that men would engage in unethical behaviors more frequently (i.e., prediction) than women, $\beta = 0.15$, SE = 0.01, 95% CI [0.13, 0.18], $t = 12.28$, $p < .001$. For ethical behaviors, participants' estimates were higher overall than for unethical behaviors ($p < .001$), but the gender difference was reversed: Women were expected to engage in ethical behaviors more frequently than men, $\beta = -0.33$, SE = 0.02, 95% CI [−0.36, −0.29], $t = -18.79$, $p < .001$. In other words, people expected women to be more ethical than men, and men to be more unethical than women.

To examine overestimation or underestimation in participants' predictions, we looked at bias scores, which are the difference between predicted and actual behavior frequencies (see Figure 2). Men's negative behaviors were overestimated relative to women's, $\beta = 0.09$, SE = 0.05, 95% CI [0.00, 0.18], $t = 1.99$, $p = .047$, while bias for positive behaviors was small and did not differ significantly between men and women, $\beta = -0.05$, SE = 0.06, 95% CI [−0.17, 0.08], $t = -0.71$, $p = .475$.

Differences in Unethical Behavior Participants predicted that men would report engaging in unethical behavior 47.2% of the time, which differed significantly from their actual rates of reporting at 35.2% (see Figure 1b). Participants predicted that women would report unethical behaviors 43.0% of the time, but only actually did so 30.9% of the time (see Figure 1b). In comparing the actual self-reported unethical behaviors, we found, as predicted, there was no statistically significant difference in the rates of unethical behavior for men compared to women, OR = 1.31, SE = 0.17, 95% CI [exp(-0.07), exp(0.60)] = [0.93, 1.82], $z = 1.58$, $p = .115$. Although these predictions mirrored Study 1, this set of behaviors did show a greater gender difference, meaning that participants were miscalibrated overall, but they predicted a gender gap that more closely matched the actual gap.

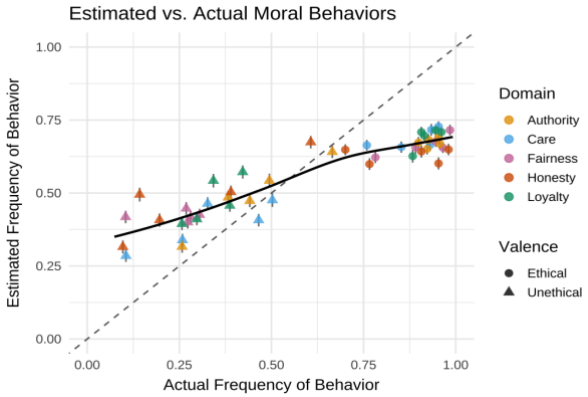

Figure 2: The relationship between empirical rates (x-axis) and predicted rates (y-axis). The dashed diagonal line represents $y = x$.

Differences in Ethical Behavior For ethical behaviors, participants underestimated engagement by more than twice the magnitude of their estimations for negative behaviors. On average, participants underestimated men's ethical behaviors by 24.2%, and women's ethical behaviors by 22.6% (see Figure 1b). Compared to the actual self-reported ethical behaviors, women reported more engagement in ethical behaviors than men, OR = 0.58, SE = 0.25, 95% CI [0.36, 0.96], $z = -2.14$, $p = .033$. While overall estimates of ethical behavior were lower than the empirical rates, participants again predicted a gender difference that aligned closely with the actual differences in behaviors.

Exploring participant gender differences We also set out to test whether accuracy of estimating rates of ethical and unethical behavior for men and women differed based on participant gender. To test this, we ran a linear mixed effects model with moral valence, target gender, participant gender, and their interaction term included as fixed effects and with by-participant random intercepts included. Results suggested participant gender did not influence prediction estimates, $\beta = -0.02$, $p = .775$. Additionally, we found no significant interactions between participant gender and valence, $\beta = 0.01$, $p = .914$.

Interim Discussion

Overall, Study 1 and 2 suggest that both men and women overestimate the frequency of unethical behavior and underestimate the frequency of ethical behavior. This cynicism about the morality of others is heightened when people are making estimates about men. One possibility is that these inaccurate predictions may stem from beliefs about differences in what behaviors men and women would be willing to report, rather than beliefs about how men and women respondents actually behave. Further, while we took care to select a range of unethical and ethical behaviors that people plausibly could have engaged in by selecting moderately common behaviors from our norming study, it is still possible that the rates that people are reporting are affected by each participant's opportunity to engage in the

behavior in question. For example, a participant may respond No to having ingested an illegal drug not because they actively declined to do so, but because they have never been placed in that situation to begin with. Differences in situational affordances such as these may explain some of the differences we observe between men and women, or lack thereof. In Study 3, we sought to test whether similar patterns from our first two studies emerge when looking at rates of hypothetical scenarios where we could hold opportunity constant and minimize concerns about reporting biases. If similar patterns of moral cynicism emerge in this hypothetical context, this would suggest that the effects observed in Studies 1 and 2 reflect biases in the rates of others' moral behavior, rather than artifacts of measurement or situational constraints.

Study 3

In Study 3, we presented participants with hypothetical decisions that involved performing questionable ethical behavior to achieve some sort of material gain. We focused on unethical behavior (vs. ethical) to directly test whether people similarly overestimate others' willingness to engage in wrongdoing when situational opportunity is held constant. We again tested whether participants overestimated gender differences in unethical behavior. We predicted overall cynicism, and that participants would overestimate the rates of unethical behavior for men to a greater extent than women.

Method

We recruited a sample of 282 participants from Amazon's Mechanical Turk online labor market (MTurk) (44% men, 56% women, M_{age} = 38 years old, *range*: 20–73 years). As an incentive for thoughtful and accurate responding, we notified participants up front that the most accurate participant in the study would receive a $25 bonus payment (accuracy was again defined as mean squared error).

Participants responded to six hypothetical decisions adapted from the Unethical Business Decisions scale (Ashton and Lee, 2008). In each scenario participants were asked if they would engage in an ethically questionable behavior to achieve material gain. For example, one scenario read as follows:

S*uppose that you are an executive for a major forest products company. In recent years, your timber production (and profits) have been reduced somewhat by new conservation laws. These laws are intended to preserve some forests for logging in future generations, but also to preserve the natural environment by securing the habitats of various species and allowing "greenhouse gases" to be removed from the atmosphere. Despite the new laws, it is very clear that the government has neither the resources nor the political will to enforce them. In fact, it is now obvious to you that your company could get away with cutting down several times more timber than the law allows. Extracting the excess timber would be extremely profitable for your company (and for your stock options), because the wood is of excellent quality, and recent forest preservation efforts have reduced worldwide supply, driving prices way up. Would you recommend that your company cut down timber beyond the legal amount?*

Each scenario was shown on a separate page, and participants responded Yes or No. On the same page, participants estimated the percentage of men and women who would say Yes to that scenario. We counterbalanced across participants whether participants first answered for themselves and then provided predictions for men and women, or made predictions first and then answered for themselves. We also counterbalanced across participants whether they first provided predictions for men and then women, or vice versa. Scenario order was also randomized for each participant.

Results

Real and Predicted Gender Differences Overall, participants expected men to be more willing than women to engage in unethical behavior. On average, participants predicted that men would be 18.1 percentage points more likely to engage in questionable behavior (see Figure 1c). We ran a linear mixed-effects model with random intercepts which showed this gap was significantly different from zero, $\beta = 0.70$, SE = 0.02, 95% CI [0.65, 0.75], $t = 28.4$, $p < .001$, indicating a clear expected gender difference. Again, we did not find a significant effect of participant gender on prediction estimates, $\beta = -0.08$, $p = .300$. However, we did find a significant interaction between participant gender and gender being rated, $\beta = -0.15$, SE = 0.05, 95% CI [−0.25, −0.06], $t = -3.17$, $p = .002$, indicating that the magnitude of the gender difference in predictions were smaller among estimates from men than women.

We observed that men were 9.2 percentage points more likely than women to report a willingness to engage in unethical behavior for material gain (see Figure 1c). This gender gap was significantly different from a null hypothesis of no gender difference ($p < .001$). We next tested whether predictions reliably deviated from reality by calculating calibration scores similar to Study 1 and 2. We again found significant overestimation of gender differences in ethical behavior ($p < .001$). On average, participants overestimated the difference between men and women by a factor of two.

We also examined whether participants were miscalibrated about men, women, or both. On average participants overestimated unethical behavior of men by 13.3 percentage points ($p < .001$) and overestimated unethical behavior of women by 4.5 percentage points ($p = 0.211$). Like Study 1, there was a general tendency to overestimate unethical behavior, but especially so for men.

Judgment Discrimination

To test for judgment discrimination we examined the correlation between predicted and actual reporting rates across participants and across questions. In general, judgment discrimination was poor. Predictions of men's response rates weakly corresponded to actual men's response rates ($r = 0.19$, $p < .001$), as did predictions of women's response rates to

actual women's response rates ($r = 0.14$, $p < .001$). Similar to Study 1, the correlation between predicted and actual gender differences was lower than that observed for prediction of men or predictions of women ($r = 0.12$, $p < .001$).

General Discussion

Across three studies we find evidence of widespread moral cynicism concerning the behaviors of others. People assume more wrongdoing and less virtue than actually exists, and predictions are shaped in part by gender expectations. In Study 1, we found that people overestimate the rates of unethical behaviors overall, but do so to a greater extent for men. In Study 2, we replicated this finding in a US representative sample. We also found that similar patterns emerge for ethical behaviors—people underestimate rates of ethical behavior, again doing so to a greater extent for men than for women. In Study 3, we test whether similar patterns emerge in hypothetical settings, finding that they do. Overall, these results suggest a miscalibration between what people expect from others' moral behaviors and what is actually true about them.

Forming accurate beliefs about the world around oneself is an important skill. Our results fit well with other research suggesting that people make systematic errors when estimating the incidence of various socially-relevant categories (Ahler & Sood, 2018). People are broadly miscalibrated with regards to others' moral character, and these biases are especially pronounced when making inferences about men. One possibility for this effect could be due to the availability heuristic (Tversky & Kahneman, 1973) that may be shaped by the aforementioned overrepresentation of men in the media (Daalmans et al., 2017). In this way, moral cynicism is not just a systematic bias overestimating the wrongdoing of others, but rather one that is more asymmetrical and structured by gendered expectations.

Importantly, miscalibrated beliefs can have downstream effects on behavior. Such beliefs can increase the likelihood that people will act in accordance with those errors. And indeed, prior research has shown consistent overestimation of a bad behavior like binge drinking can increase how often people engage in that themselves (e.g., Prentice & Miller, 1993; see also Manstead et al., 1992; Young & Weerman, 2013). These findings extend this work to the moral domain. If people assume others, especially men, are immoral, are they more likely to be immoral themselves? Future research should investigate whether people higher in the belief that others are immoral are more likely to endorse those behaviors themselves.

One limitation of the present studies is that the rates of good and bad behavior we use to benchmark for accuracy are self-reported in Study 1 and 2. Prior research has suggested that the accessibility of memories of morally relevant events is asymmetric—people tend to forget the things that they've done in the past that were immoral (i.e., unethical amnesia; Kouchaki & Gino 2016; Helion et al., 2020), but not the things that were morally good. This may also serve as an explanation behind potential ceiling effects in self-reported ethical behaviors. Thus, it is possible that our overestimation of unethical behavior is merely a product of participant-driven recall errors, rather than reflecting an accurate estimate of the rates of those behaviors. However, we think there are at least two reasons that this concern should be assuaged. First, we find similar patterns when looking at hypothetical behaviors whereby participants are asked not to report their own unethical behaviors, but hypothetical rates. Second, we find similar cynicism when we have participants estimate morally good behaviors. That is, the research on motivated memory for moral events suggests that people should be more accurate when recalling morally good events. Future research could directly test the rate of actual ethical behaviors without any risk of reporting biases by measuring prosociality indirectly, such as through direct assessment of performance or behavior in a controlled setting. And yet even for ethical behaviors we see a similar pattern of results as in Study 1 and Study 3. Together, we take the combination of Study 1-3 as evidence to suggest that across samples, stimuli, and the limitations of human memory, the errors presented here reflect a combination of cynicism and gendered stereotypes.

A limitation specific to our third study is the type of trade-off included in our hypothetical scenarios. With all scenarios presenting material gain as the only benefit, the self-reported results may align more closely with traditionally masculine stereotypes. Future research should investigate whether similar effects emerge when the trade-off value is presented differently, such as involving familial well-being.

We set out to examine whether people can accurately estimate rates of moral behavior. In particular, we were interested in whether beliefs about the moral goodness of women (e.g., Wylie et al., 2025) would bias estimates of individual moral behaviors for men. And indeed, results from the three studies presented here suggest that people expect men to be morally worse than women. In Study 1, we find that this estimate is unfounded—men and women report similar rates of unethical behavior. However, in Study 2, we find that, though the magnitude of the effect is not accurate, the direction is. Men report more unethical behaviors and fewer ethical behaviors than women. It is possible that, though people have a tendency to view others through a cynical lens, they may be somewhat correct in their expectations about men. Future research should further examine how gender influences moral behavior, across a range of methods.

People rely on beliefs about others' moral character to navigate social life. But these beliefs are distorted: Expectations and biases shape how people imagine the moral world around them. We found that people are broadly pessimistic about the moral behavior of others—an effect which gets stronger when people think about men (vs. women). Together, these results demonstrate how distorted expectations, shaped in part by gender stereotypes and moral cynicism, may influence how people perceive the behaviors of others in everyday life.

References

Ahler, D. J., & Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics, 80*(3), 964-981.

Ashton, M. C. and Lee, K. (2008). The prediction of honesty–humility-related criteria by the hexaco and five-factor models of personality. *Journal of Research in Personality, 42*(5), 1216–1228.

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.

Bosworth, S. J. (2017). The importance of higher-order beliefs to successful coordination. *Experimental Economics, 20*(1), 237-258.

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). Academic Press.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1–3.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology* (Vol. 24). Academic Press.

Croson, R., Handy, F., & Shang, J. (2009). Keeping up with the Joneses: The relationship of perceived descriptive social norms, social information, and charitable giving. *Nonprofit Management and Leadership, 19*(4), 467-489.

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*(5), 366-378.

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., & 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.

Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking: reactive egoism in social interaction. *Journal of personality and social psychology, 91*(5), 872.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

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.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American psychologist, 56*(2), 109.

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., 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.

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.

Guay, B., Marghetis, T., Wong, C., & Landy, D. (2025). Quirks of cognition explain why we dramatically overestimate the size of minority groups. *Proceedings of the National Academy of Sciences, 122*(14), e2413064122.

Haines, M., & Spear, S. F. (1996). Changing the perception of the norm: A strategy to decrease binge drinking among college students. *Journal of American College Health, 45*(3), 134-140.

Helion, C., Helzer, E. G., Kim, S., & Pizarro, D. A. (2020). Asymmetric memory for harming versus being harmed. *Journal of experimental psychology: general, 149*(5), 889.

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.

Johnson-Laird, P. N. (2004). The history of mental models. *In Psychology of reasoning*. Psychology Press.

Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences, 113*(22), 6166-6171.

Kruger, J., & Gilovich, T. (1999). " Naive cynicism" in everyday theories of responsibility assessment: On biased assumptions of bias. *Journal of Personality and Social Psychology, 76*(5), 743.

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.

Lei, R. F., Leshin, R. A., Moty, K., Foster-Hanson, E., & Rhodes, M. (2022). How race and gender shape the development of social prototypes in the United States. *Journal of Experimental Psychology: General, 151*(8), 1956.

LeVier, T., Wylie, J. (2026). The influence of gender on accuracy in estimating moral behavior [OSF project]. OSF. https://osf.io/hpwna/overview?view_only=6e882d8121ff4a32af61bc3567170868

Manstead, A. S., Parker, D., Stradling, S. G., Reason, J. T., & Baxter, J. S. (1992). Perceived consensus in estimates of the prevalence of driving errors and violations. *Journal of Applied Social Psychology, 22*(7), 509-530.

Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles, 75*(9), 409-421.

Murphy, A. H. (1973). A new vector partition of the probability score. J*ournal of Applied Meteorology, 12*(4), 595–600.

Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. *Journal of personality and social psychology, 64*(2), 243.

Prentice, D. A., & Miller, D. T. (1996). Pluralistic ignorance and the perpetuation of social norms by unwitting actors. In *Advances in experimental social psychology* (Vol. 28). Academic Press.

Rhodes, M., & Baron, A. (2019). The development of social categorization. *Annual review of developmental psychology, 1*(1), 359-386.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. Journal of Social Issues, 57(4), 743–762.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science, 18*(5), 429-434.

Shtulman, A., & Phillips, J. (2018). Differentiating "could" from "should": Developmental changes in modal cognition. *Journal of Experimental Child Psychology, 165*, 161-182.

Strohminger, N. (2018). Identity is essentially moral. *Atlas of moral psychology*.

Stuhler, O. (2024). The gender agency gap in fiction writing (1850 to 2010). *Proceedings of the National Academy of Sciences, 121*(29), e2319514121.

Tetlock, P. E. (2017). *Expert Political Judgment: How Good Is It? How Can We Know?-New Edition.* Princeton University Press.

Torgler, B., & Valev, N. T. (2006). Women and illegal activities: Gender differences and women's willingness to comply over time. *Andrew Young School of Policy Studies Research Paper,* (06-56).

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive psychology, 5(2), 207-232.

Wylie, J., Gantman, A., Mende-Siedlecki, P., Singer, H., Young, L., & Bloom, P. (2025). Moral Goodness is Gendered. *(under review).*

Young, J. T., & Weerman, F. M. (2013). Delinquency as a consequence of misperception: Overestimation of friends' delinquent behavior and mechanisms of social influence. *Social Problems, 60*(3), 334-356.