**Running Head: Kindness Bias**

**A Bias Toward Kindness Goals in Performance Feedback to Women (vs. Men)**

**Word Count: 10,835**

**Abstract**

While research has documented positivity biases in workplace feedback to women versus men, this phenomenon is not fully understood. We take a motivational perspective, theorizing that the gender stereotype of warmth shapes feedback givers’ goals, amplifying the importance placed on kindness when giving critical feedback to a woman versus a man. We found support for this hypothesis in a survey of professionals giving real developmental feedback (Study 1, N= 4,842 raters evaluating N=423 individuals) and 5 experiments with MBA students, lab participants, and managers (Studies 2-5, N=1,589). Across studies, people prioritized the goal of kindness more when they gave or anticipated giving critical feedback to a woman vs. a man. Studies 1, 3, and 5 suggest that this kindness bias relates to gendered positivity biases, and Studies 4a-4b tested potential mechanisms and supported an indirect effect through warmth. We discuss implications for the study of motivation and workplace gender bias.

KEYWORDS: Feedback; Gender bias; Goals; Kindness

**A Bias Toward Kindness Goals in Performance Feedback to Women (vs. Men)**

**Introduction**

Past work (within the gender binary) suggests that women receive differentially valanced feedback, compared to men, about their workplace performance. Critical feedback to women has been described as reflecting a “positivity bias,” (Fisher, 1979; Fisher, 1984; Waung & Highhouse, 1997), whereby evaluators adjust developmental feedback to include less negative content. Women employees self-report receiving more positive feedback than male colleagues (King et al., 2012; also see Lean in & McKinsey & Company, 2015) and are given inflated feedback compared to men when underperforming (Jampol & Zayas, 2020). Although such positivity biases may seemingly serve to benefit women, inflated feedback prevents learning from mistakes, reducing the likelihood of improvement and advancement over time, especially in situations where critical feedback is needed for improvement (Barney, 1986; Correll et al., 2020; London, 2003; Sørensen & Sorensen, 2002). Further, positive feedback becomes less meaningful and prognostic for women than men: women are less likely than men who receive similarly positive feedback to receive prestigious job assignments, raises, or promotions (Biernat et al., 2012; Vescio et al., 2005).

The processes that drive gender biases in the positivity of performance feedback remain underexplored. While past work has focused on identifying gender biases in the content or accuracy of feedback (Biernat et al., 2012; Jampol & Zayas, 2020; Vescio et al., 2005), we seek to advance the field’s understanding of gendered feedback biases by starting to explain why these differences arise. Indeed, similar positivity biases have been documented for critical feedback to racially minoritized versus majority groups (i.e., concerns about appearing racist; Harber, 1998; Harber et al., 2010). This work in the domain of race suggests that it is possible to unpack the psychological drivers of positivity biases in developmental feedback, as we hope to do in the context of gender. This is a crucial next step because understanding the underlying processes that shape evaluators’ differentially positive feedback to women and men will facilitate future research to design and test interventions that encourage equal treatment.

**What drives positivity biases?**

**Hypothesized Process: Stereotype Content.** Our core theory draws together the study of stereotyping and motivation. We propose that a feedback recipient’s gender activates stereotype content knowledge in the mind of evaluators, cognitive activation which spills over to shape how they prioritize feedback goals as they enter feedback-giving processes. Stereotypes characterize women as higher in warmth but lower in competence relative to men (Eagly et al., 2020; Fiske et al., 2002; Rudman et al., 2012). Past work has shown that groups stereotyped as high in warmth elicit consideration, care, concern, and sympathy from observers especially when members of these identity groups need support (Cuddy et al., 2008; Fiske et al., 2002; Glick & Fiske, 1996). In other words, the gendered activation of warmth stereotypes in perceivers’ minds can spill over to foster support-oriented behavioral intentions toward a target (Cuddy et al., 2007). In this research, we theorize that the gendered activation of warmth stereotypes in perceivers’ minds can also spill over to shift the feedback goal priorities people exhibit when giving developmental feedback.

Our work does not propose a new feedback-giving goal, but instead suggests that the feedback recipient’s gender shapes the priority placed on already-active motivational processes among feedback givers (which is why we describe our theory in terms of spillover rather than spreading activation). During feedback giving, people have multiple goals. While people giving developmental or critical feedback are especially motivated to be candid about where someone needs to improve, another salient goal that arises simultaneously is kindness (Annett, 1969; Lupoli et al., 2017; Wiltermuth et al., 2015). In other words, when giving feedback people are already motivated to offer both candor and kindness. Advancing the study of the Stereotype Content Model by integrating it with the study of goal priorities during developmental feedback-giving, we theorize that the stereotypic association of women with warmth can spill over to reprioritize the feedback goals that feedback-givers hold. Specifically, we hypothesize that when people anticipate giving feedback to a woman, rather than a man, they amplify the importance of kindness as a feedback goal. Because goals drive behavior (Locke & Latham, 2002), we expect kindness goals to motivate feedback givers to exhibit greater positivity in feedback to women. We test our argument that this is driven by warmth in two ways, directly by assessing perceptions of warmth and indirectly by testing whether people see kindness as more helpful for women.

In addition to the hypothesized process described above, we investigate multiple possible alternatives.

**Alternative 1: Competence.** Our theory focuses on warmth as the driver of gendered positivity biases for two reasons. First, the Stereotype Content Model gives primacy to warmth because it characterizes behavioral intentions, and describes competence as the secondary dimension because it characterizes others’ capability to carry out those intentions (which necessarily follows in importance and order from assessing intentions; Cuddy et al., 2008). Second, the behavioral intentions that follow from higher warmth perceptions are active facilitation or helping, which overlap cognitively and in agency with giving developmental feedback, while the behavioral intentions that follow from lower competence are passive inaction or neglect, which would invoke a lack of motivation to give developmental feedback, rather than heightening a feedback goal (Cuddy et al., 2007; 2008). Following this theory, we did not expect gendered differences in perceptions of competence to explain positivity biases (Cuddy et al., 2008), but nonetheless we test this directly as well as indirectly by evaluating whether the proposed effects emerge in feedback toward both high and lower performers across studies.

**Alternative 2: Hostile Sexism.** While people can identify where groups fall along the dimensions of warmth and competence by virtue of their stereotype *knowledge*, this does not capture their *affective negativity*, a type of prejudice known as sexism. It is possible that positivity biases arise out of people’s dislike of women and a resulting desire to sabotage women’s workplace outcomes. While we did not predict this, we evaluate this alternative. Given that research has shown that women exhibit less hostile sexism than men (Glick & Fiske, 1996), we considered testing for differences based on gender of the feedback-giver an indirect way of testing for hostile intent toward women in the workplace. Second, we measured another, if anything more critical, feedback goal priority across studies. Candor, being honest and straightforward about flaws, failures, or areas for improvement, is the most important goal when giving feedback (Bergsieker et al., 2012; Levine & Schweitzer, 2014; Lupoli et al., 2017; Tesser & Rosen, 1975), and thus we considered candor as an indicator that might pick up hostile sexism. That is, if people seek to strategically undermine women in the workplace, then they should not only inflate the priority they place on kindness – they might also reduce the priority on candor. We test this across studies, which also ensures that people do not simply inflate all ratings after reading about a woman (vs. man).

**Alternative 3: Benevolent Sexism.** Scholarship has also identified a more purportedly benevolent form of sexism whereby people paternalistically protect women, who are deemed fragile, while also keeping them “in their place” (Glick & Fiske, 1996). Benevolent sexism is another form of affective negativity toward women, though it is wrapped in the presentation of paternalistic care and concern. Perhaps inflating the priority placed on kindness as a feedback goal arises from these sexist protective instincts. Though we did not predict that benevolent sexism would account for the proposed effects, we test this in two ways. We directly assess benevolent sexism, and we also assess the sense of women as fragile in the face of negative feedback by separately measuring perceptions of negative emotion after receiving negative feedback.

**Alternative 4: Shifting Standards.** The extensive body of research on shifting standards documents that people apply within-group referents when evaluating women and men on subjective dimensions (Biernat & Manis, 1994; Biernat & Vescio, 2002). People evaluating a woman’s (man’s) poor performance judge her (him) relative to other women (men) - the standard of judgement shifts as a function of gender. A shifting standards account of positivity biases would suggest that people are kinder to women (vs. men) because her poor performance does not seem as bad against the (gender-relative) standard of judgement. To address this possibility, later studies always specify performance on an objective, numeric evaluation, leaving less room for shifting standards (Jampol et al., 2020). Secondly, we assess how much the low performance violates expectations, because a shifting standards account would suggest that people exhibit more surprise at a poorly performing man than woman, given the different standards of judgment.

**Alternative 5: Concerns about appearing biased.** Given past research on race and positivity biases (Harber, 1998; Harber et al., 2010), we tested for concerns about appearing biased against women. We did not predict this because people tend to exhibit less social evaluative concern in the context of gender than in the context of race (Apfelbaum et al., 2008), but we measured internal and external motivation to control prejudice in the context of giving feedback to women at work in order to assess the possibility. In case people’s concern about appearing biased is more generalized than would be captured by the these constructs, we also measured general discomfort with giving feedback.

**Alternative 6: Stereotypes about Men.** It is also possible that positivity biases are instead driven by men-specific stereotype content, in particular the stereotype of men as agentic and as “fighters.” If people expect more disagreement when giving feedback to a man than a woman, they may exhibit less anticipatory kindness. We tested whether stereotypes about men as disagreeable play a role by measuring feedback givers’ fear of conflict with the recipient, though we did not predict differences.

**Overview of Studies**

The core contribution of this research is to theoretically advance the study of gender bias in the workplace by considering how gender shapes evaluators’ prioritization of goals during developmental feedback – goals which may yield differentially positive feedback to women versus men. Our hypothesis was that people view kindness as a more important goal when giving feedback to a woman versus a man, due to the association of women with warmth. We further hypothesized that feedback givers’ kindness goals would be associated with more positive feedback content. Six studies test these hypotheses. First, we test for real-world evidence of gender-biased goal priorities and feedback positivity in a correlational study of actual performance feedback (Study 1). Next, we shifted to an experimental approach. We tested whether MBA students (Study 2), lab participants (Study 3, 4b), and professional managers (Studies 4a and 5) would prioritize kindness as a feedback goal more when anticipating giving developmental feedback to a woman versus a man. Across studies, we explored whether gender of the feedback recipient would also shape how much people prioritize the feedback goal of candor. We also tested the process by which we theorize these effects arise, and the alternate mechanisms described above. In Studies 4a-b we test our hypothesized process directly by measuring warmth perceptions, and in Study 5 we assess it indirectly by evaluating whether kindness is seen as more helpful to women than men. To provide a more reliable estimate of the direct effect of gender on the prioritization of kindness as a developmental feedback goal, we also present a mini-meta-analysis for the core hypothesized effect of gender on the priority placed on kindness as a goal.

**Study 1**

**Methods**

In this and all studies, we report all participants recruited, conditions, measures, and any exclusions. The studies were run in a single wave and data was only analyzed after data collection closed. Due to the EU/UK GDPR data protection and privacy law, the data from this study cannot be shared publicly. Data and analysis syntax (including codebook information) for all other studies are available on OSF (redacted).

**Participants**

Study 1 uses a field dataset of ratings given to Masters in Business Administration (MBA) students from their former co-workers, originally collected for teaching purposes. The data exists at 2 levels: 423 MBA students (259 men, 160 women, 4 unknown) who were evaluated by 4,842[[1]](#footnote-1) of their previous coworkers. The MBA students in this sample were generally good performers in their previous jobs (the business school is ranked among the top 5 global business schools) who will have received positive feedback from others in their organizations (e.g., references from co-workers are a key part of applications for admission to the school). The school is located in the U.K. and students represent over 60 nationalities. Because at least some raters typically come from students’ nation of origin, we can assume that raters were similarly international (though no rater demographics were collected in order to ensure rater anonymity).

**Procedure**

**360 ratings.** A 360 survey is a common feedback-giving practice in organizations, whereby a third party (in this case a contractor) solicits feedback about an individual from current or previous co-workers who represent a ‘360’ view (i.e., supervisors, peers, and subordinates). In this case, it was clearly stated to all parties that the 360 would be fully anonymized – students would only receive summary information that aggregated across enough raters to protect anonymity. This minimizes the possibility of self-presentational concerns shaping evaluators’ ratings and feedback.

Before the start of their MBA, all incoming students received a link to a ‘360 survey’. Students were instructed to nominate people to offer feedback about their performance. The instructions encouraged students to list as many people as possible from supervisors, mentors, peers, and subordinates, suggesting 20 people would be a fair number but encouraging them to list as many as possible. Once students entered raters’ contact details, raters received an email requesting that they complete an anonymous survey and provide quantitative and qualitative feedback. This survey was part of incoming students’ coursework, and an academic requirement. Students received the feedback provided, though it was presented in aggregate (for quantitative data) and anonymized (for all data). Prior to completing the survey and nominations, students provided consent approving the use of their data for academic research and raters for participation.

**Feedback Goal Priority.** Raters first responded to a routine set of ‘360 survey’ items, which were for academic teaching purposes only and were not accessed or analyzed by the researchers. At the end of the survey, participants answered two questions about their feedback goal priorities (due to restrictions on survey length, we were only able to include single-item measures). To measure evaluators’ kindness goal priority, our focal outcome variable, we asked raters, “Thinking about the ratings you just made, how much were you trying to be kind, considerate, and compassionate as you communicated about what she/he needs to improve?” To measure the goal priority of candor, raters responded to the item, “Thinking about the ratings you just made, how much were you trying to be straightforward, blunt, and direct as you communicated about what she/he needs to improve?” (scales: 1 “Not at All” – 7 “Extremely”).

**Feedback Style.** We assessed differences in feedback style in two ways. First, to test whether people have explicit knowledge of any positivity bias they exhibit, we asked them to rate “How positive (i.e., versus negative) was the feedback you gave, overall?” (scale: 1 “Mostly Negative” - 7 “Mostly Positive”). Secondly, to test for evidence of a positivity bias in the written feedback, we processed all of the raters’ responses to the five qualitative questions in the survey (“What does this person do well?” “What are the two most important values or beliefs that this manager holds about work and the way the organisation should operate?” “In what situation is this person at their very best?” “If you were asked to advise this manager on how he/she could be more effective as a leader what would you suggest?” “What career/personal development needs do you think this manager has?”) through the LIWC text analysis program (Pennebacker et al., 2015), which uses normalized dictionaries to quantify the tone of qualitative text. We pre-identified the outputs provided by LIWC most relevant to positivity biases: emotional tone (i.e., an algorithmic summary variable capturing tone positivity; see Cohn, Mehl, & Pennebaker, 2004), positive emotion (i.e., words such as love, nice, sweet), and negative emotion (i.e., words such as hurt, ugly, nasty).

**Results**

To test our hypotheses, we conducted multilevel models with restricted maximum likelihood estimation using feedback recipient as the subject identifier (Rabe-Hesketh & Skrondal, 2012). Prior work shows that feedback recipients’ experience and the type of relationship between the giver and recipient influence feedback content (Harris & Schaubroek, 1988). In all analyses, therefore, we controlled for the relationship type between rater and recipient (supervisor, direct report, peer, or customer) and age of the feedback recipient as a proxy for experience. Controlling for experience was particularly important in this dataset because, consistent with the institution where the data were collected, women MBAs (M=28.25, SD=2.14) were significantly younger (i.e., less experienced) than men (M=28.97, SD=2.43), F(1, 415)=9.45, *p*=.002, ηp2=.02.

**Feedback Goal Priority.** We conducted a two-level random intercept model to examine the influence of feedback recipient gender on raters’ self-reported feedback goal priorities. In line with our hypothesis, participants indicated that they prioritized the goal of kindness more in feedback to women than men, *b* = .11*, z* = 2.11, *p* = .035, 95% CI [0.007, 0.21], see Table 1. The exploratory analysis of how recipient gender influenced raters’ priority on candor indicated that evaluators did not significantly differ in the goal priority placed on candor for men versus women, *b* = -.04*, z* = 1.65, *p* = .099, see Table 1, though the trend suggests marginally more importance on candor to men than women.

**Feedback Style.** Evaluators did not self-report giving more positive feedback to women than men, *b*=-0.02*, z*=0.78, *p*=.44, 95% CI [-0.08, 0.04]. However, analysis of their qualitative answers revealed that, as predicted, their comments had a significantly more positive tone, *b*=3.31*, z*=2.88, *p*=.004, 95% CI [1.06, 5.58], included a higher percentage of positive emotion words, *b*=0.54*, z*=2.93, *p*=.003, 95% CI [0.18, 0.90], and included a lower percentage of negative emotion words, *b*=-0.20*, z*=2.36, *p*=.02, 95% CI [-.37, -.03], in feedback to women than men, see Table 2.

**Indirect Effects.** We theorized that recipient gender shapes feedback-givers’ motivation to be kind, which shapes the positivity of actual feedback. We therefore explored whether kindness goal priority statistically mediated the relationships between recipient gender and feedback tone, positive emotion, and negative emotion using 2-1-1 multilevel mediation analyses (see Tingley et al., 2014), again controlling for age and relationship type and using bias corrected bootstrap confidence intervals (Preacher & Hayes, 2008). Contrary to our predictions, we did not find support for feedback tone or for negative emotions (indirect effect estimates<.02, bias-corrected 95% CIs crossed 0, 1000 repetitions). As predicted, however, a significant indirect effect of gender on positive emotions used in the feedback through kindness intentions emerged (indirect effect estimate= .035, 95% CI [.00013, 0.13]). While supported, we interpret this result cautiously given that the confidence interval approached (though it did not cross) zero.

**Study 1 Summary**

Study 1 offers evidence from the field that professionals focus more on kindness as a goal when giving developmental feedback to women than men. Further, it provides initial evidence that the higher priority placed on kindness goals for women can in turn increase feedback positivity, through positive emotion-related words. While this evidence from consequential feedback between co-workers in the field is compelling, it leaves open the question of causality and also was not conclusive in terms of the results on candor, an indirect indicator of whether hostile sexism may be at play. Further, the feedback recipients in this sample were high performers, given that they gained entry into a top business school. Therefore, in Studies 2-4, we turned to controlled experiments to explore how recipient gender influences evaluators’ feedback goals when giving feedback to identical, low-performing men versus women.

**Study 2**

**Methods**

**Participants**

Three-hundred-forty-nine international MBA students at a UK university (222 men, 126 women, 1 indicating ‘other’; age and racial/national background not recorded) participated. Our sample was a convenience sample – the sample size was determined by the number of students who completed the survey and provided informed consent – though appropriate as MBAs are experienced with feedback-giving in formal settings.

**Procedure**

Participants completed informed consent and a series of unrelated measures collected for teaching purposes, amid which the following were included.

**Scenario.** Participants read: “Imagine that you are in your next job and, today, you are doing performance review meetings with your four direct reports. Although all four members of your team are performing above the minimum basic standard, one is clearly lagging behind the others in performance so you are thinking deeply about how to manage this performance review.”

**Gender Manipulation.** The employee’s name manipulated gender (Eagly & Wood, 1982; Moss-Racusin et al., 2012; Swim et al., 1989; also see SOM). Participants either read that the lowest performing team member was “Andrew” (N = 173) or “Sarah” (N = 176). Participants were told, “you must give [Andrew/Sarah] a clear indication of his/her low performance” and that much of the meeting would be focused on the employee’s improvement.

**Feedback Goal Priority**. Participants were asked to “Think about the goals you would have going into your one on one performance feedback meeting with [Andrew/Sarah].” Participants indicated “How much do you think each of the following would be a TOP PRIORITY for you when giving feedback to [Andrew/Sarah]?” (scale: 1 “*Not at all*” - 6 “*Extremely”*). They responded to six items representing the two constructs of interest: Kind goals (“To be sympathetic and compassionate,” “kind and considerate,” “avoid hurting feelings”; alpha = .56) and Candid goals (“To be direct and to the point,” “completely straightforward and candid,” “avoid being sentimental”; alpha = .59). As expected, the items loaded onto two factors and so means of each were calculated and used in all analyses (see SOM for confirmatory factor analysis details).

*Exploratory Measure.* As an exploratory measure, participants were asked to write about any goals that were not listed, which they might have going into the feedback meeting. This data was not formally coded or analyzed because, upon initial review, we could see that responses had extremely high variance and no clear patterns or repeated themes emerged.

Participants finally indicated their gender and were debriefed.

**Results**

Means, standard deviations, and correlations are presented in Table 3 for the main analyses.

**Feedback Goal Priority**. Our core hypothesis is about gender differences in feedback priority for kindness. To directly test this, our a priori analysis plan was to conduct independent samples t-tests to examine whether participants would prioritize kindness goals more for the woman employee than for the man. As predicted, there was a gender difference in how much kindness was prioritized as a goal. MBA students prioritized the goal of kindness significantly more for Sarah than for Andrew, *t*(347) = -2.39, *p* = .02, Cohen’s *d* = -.26, 95% CI [-0.46, -0.05]. Participants showed no difference in goal priority for candor to Sarah and Andrew, *t*(347), *p* = .24 (see Figure 1), suggesting hostile sexism is not at play.

**Evaluator gender.** We thought it important to explore whether women and men showed this pattern of effects to an equal degree, which as noted above could speak to the potential of hostile sexism driving any effects. Therefore, we conducted exploratory analyses of whether participant gender influenced ratings. Because only 1 person identified as 3rd gender, this analysis focuses on the gender binary. There was no participant gender by condition interaction on kindness goals, *F*(1, 347) = 0.45, *p* = .50, *np2* = .001. Controlling for participant gender, the main effect of condition on kindness goals (described above) was still significant *F*(1, 347) = 5.25, *p* = .02, *np2* = .02. Because we do not find consistent evaluator gender effects in any studies, we do not discuss them further (see SOM).

**Study 2 Summary**

Feedback givers prioritized kindness as a goal more when giving developmental feedback to women than men. Both men and women exhibited this differential tendency equally, indirectly suggesting hostile sexism is not at play. However, evaluators may have had different impressions of what constituted “lagging performance” for the woman versus the man (e.g. “shifting standards,” Biernat & Manis, 1994). To minimize this possibility, the next study provided objective information about performance.

**Study 3**

**Method**

**Participants**

A total of 94 participants from a UK university and surrounding community (61 female; 10 over 50 years of age, 16 between 30 and 50, 68 between 18 & 30; 44 White, 25 South Asian; 10 East Asian, 6 Black, 1 Arab, 1 Hispanic, and 7 Mixed or Other) took part in a laboratory experiment in return for £10 GBP. No participants were excluded. This was a convenience sample – our study was added to an already running (unrelated) lab study that did not take up the full participant time slot, and therefore our sample size was determined by how many people participated in that study.

**Procedure**

After providing informed consent, participants were asked to imagine themselves as a manager of a team of six people in an advertising agency. They were told that on that day they would have to give each team member performance feedback in preparation for a more formal review six months from the present.

**Manipulation.** Again, employee name manipulated gender, Sarah or Andrew. All participants read that [Andrew/Sarah] “has a 25% competency score, putting her / him in the bottom quarter of performers on the team, and at risk for being fired.”

**Manipulation Check.** Participants were asked to type the employee’s name into a text box. All participants correctly recalled the employee name.

**Feedback Goal Priority.** Participants again indicated their agreement on the items from Study 2, with one additional item for each factor, for a total of eight items (scale: 1 “*Not at all*” - 6 “*Extremely”*): Kindness goal priority items ‘to be kind and considerate’, ‘to avoid hurt feelings or a bad reaction’, ‘to maintain a positive relationship’, and ‘to be sympathetic and compassionate’ (alpha = .82); Candid goal priority items: ‘to be fair and unbiased’, ‘to be completely straightforward and candid’, ‘to be firm and unsentimental’, and to ‘be direct and to the point’ (alpha = .61). Order of item presentation was randomized.

**Feedback Style.** We sought to explore the link between the priority placed on kindness and the positivity evaluators anticipated exhibiting in their feedback. Thus, we included exploratory measures that we hoped would capture three different ways in which feedback can come to be inflated – if it is overly positive, lacks critique, or focuses more on strengths than weaknesses. We asked participants to rate on bipolar scales their anticipated ‘feedback valence’ (1 “*Extremely negative feedback*,” – 6 “*Extremely positive feedback*”*)*, ‘feedback straightforwardness’ (1 “*Extremely straightforward/All critical feedback*,” – 6 “*Not at all straightforward/No critical feedback*,” reverse coded so that higher numbers indicate more straightforwardness), and ‘feedback focus’ (1 “*Focusing on aspects of the work that need a lot of improvement*,” – “*Focusing on aspects of the work that are already going well*,”*)*.

Finally, participants indicated their gender, age, education level, and political orientation and were debriefed for the study session (see SOM for additional exploratory measures and results).

**Results**

Means, standard deviations, and the correlations for key variables are in Table 4.

**Feedback Goal Priority.** An independent samples t-test assessed the focal hypothesis. Replicating both previous studies, and as predicted, participants indicated that they would prioritize kindness goals more during feedback to Sarah than to Andrew, *t*(92) = 1.95, *p* = .05, *d* = 0.41, 95% CI [0.001, 0.82]. Also consistent with previous studies, people did not prioritize candid goals significantly more for Sarah than Andrew, *t*(92) = 0.93, *p* = .35 (see Figure 1).

**Feedback Style***.* We analyzed the data using model 4 in Process (Hayes, 2012), with 20,000 bootstrapped samples, entering Gender Condition as X, Feedback Style (Valence, Straightforwardness, Focus) as Y, each in turn, and Kindness Goals as M. When participants anticipated giving feedback to a woman versus man, their higher kindness goal priority predicted greater anticipated positive feedback, indirect effect *B* = .22, *SE* = 0.12, 95% CI [.002, .49]; total effect *B* = -.26, *SE* = .20, 95% *CI* [-.66, .14]. The indirect effect was not supported for straightforward feedback, indirect effect *B* = -.15, *SE* = 0.10, 95% CI [-.4, .001], total effect *B* = -.05, *SE* = 0.21, 95% CI [-.46, .36], or for expected focus on strengths, indirect effect *B* = -.10, *SE* = 0.07, 95% CI [-0.27, 0.004], total effect *B* = -.09, *SE* = 0.21, 95% CI [-.51, .33]. These results suggest that the greater priority people place on kindness goals when anticipating giving performance feedback to a woman (vs. man) may foster feedback that reflects positivity biases, though we note that the CIs are close to zero and thus these results should be taken as suggestive.

**Study 3 Summary**

These results offer another replication of the focal effect – people rated kindness as a higher goal priority when randomly assigned to anticipate giving critical feedback to a woman, compared to a man. Their higher priority of kindness goals for the woman (vs. man) was in turn associated with reporting more anticipated positivity.

The next two studies offer pre-registered replications to test the full proposed model: feedback recipient gender differentially activates warmth, which explains the higher priority people place on kindness as a feedback goal. We also compare our proposed mechanism to a variety of alternative possibilities. These studies occurred early in the COVID-19 pandemic, which saw people working from home in record numbers. In this time, giving and receiving feedback to an employee was fundamentally different than in the previous studies, and the record disruption (physical, emotional, and economic) created a cultural moment of substantial extra attention to kindness in the workplace. Thus, while these studies sought to replicate the earlier ones, it also took place under dramatically different conditions.

**Study 4a**

**Method**

**Participants**

This was pre-registered (blinded for review). An a priori sample size calculation indicated that a total sample size of N = 506 would provide sufficient power (based on effect size d = .25 from an earlier version of the mini-meta-analysis, two-tailed, alpha error probability = .05, power = .8, and n2/n1 = 1). In order to ensure we achieved the required sample size, we oversampled to collect a panel sample of 650 participants who were currently in managerial or supervisory roles in the United States and the United Kingdom. All participants who failed one or more of the a priori exclusion criteria (failed the attention check, failed to accurately recall the employee name, or did not have managerial responsibilities), were excluded from the analysis, leaving 527 participants (296 ‘woman’, 227 ‘man’, 4 ‘nonbinary/genderqueer,’ 1 ‘trans’; mean age = 37.43, SD = 11.15; 15 ‘Black/African,’ 1 ‘Latino/Hispanic,’ 6 ‘East Asian,’ 3 ‘South East Asian,’ 31 ‘South Asian,’ 2 ‘Middle Eastern,’ 464 ‘White/Caucasian,’ 8 ‘Other,’ note that participants could choose multiple options for both gender and race identifications).

**Procedure.** Participants read the same scenario as in Study 3: they were about to give feedback to a poor performer [Andrew/Sarah] who had scored a 25% competency level.

**Manipulation Check.** Participants were asked to provide the name of the employee they had just read about.

**Feedback Goal Priority.** Participants completed the same measures as in Study 3 (kindness alpha = .74, candor alpha = .64).

**Mechanisms.**

**Warmth.** Participants completed the classic measure of warmth from Fiske and colleagues (2002), rating how much they saw the employee as warm, trustworthy, friendly, and sincere (all scales, unless otherwise noted: 1 “not at all” - 6 “extremely,” alpha = .91).

**Competence.** Participants completed the classic measure of competence from Fiske and colleagues (2002), rating how much they saw the employee as competent, capable, intelligent, and skillful (alpha = .90).

**Expectancy Violation.** If people use a lower standard of judgement for women, as the shifting standards literature might suggest, then they should be less surprised by a woman performing poorly than a man, and this could explain the effects observed in the earlier studies. To test this account, we asked participants how surprised, unexpected, and problematic they saw the employees’ poor performance to be (alpha = .69).

**Discomfort.** It was possible that people might prioritize kindness toward women because they are uncomfortable giving negative feedback to a stigmatized group. To assess this, participants rated how comfortable and awkward (reversed) they felt about giving feedback to the employee (r = .68, p < .001, from Wuang & Highhouse, 1997).

**Fear of Conflict.** Alternatively, it was also possible that people exhibit less kindness to men because they worry that men will be aggressive or disagreeable in response. To get at this, we measured how concerned participants were that the employee would disagree with the feedback (scale: 1 “very unconcerned” - 6 “very concerned”; from Wuang & Highhouse, 1997).

**Perceived emotions.** Another possibility was that participants anticipated a woman reacting more emotionally than a man, which could evoke more kindness. We measured participants’ perceptions of how the employee would feel after the feedback using the 15-item negative affect subscale of the PANAS (e.g., distressed, upset, cite; scale: 1 “very slightly or not at all” - 5 “extremely,” alpha = .93). We note that our measurement for the alternative mechanism of perceived emotions asked how the feedback recipient would feel after the feedback, which does not answer the question of whether participants’ anticipations of emotion differ prior to the feedback.

**Benevolent Sexism.** We directly measured benevolent sexism in this study using the subscale from Glick and Fiske’s (1996) Ambivalent Sexism Inventory (e.g., “Women should be cherished and protected by men.” Scale: 1 “disagree strongly” - 6 “agree strongly,” alpha = .85).

**IMS/EMS.** We measured internal (alpha = .76) and external (alpha = .81) motivation to control prejudice, adapted to specify controlling prejudice toward women during feedback, using the measure from Plant and Devine (1998).

**Results**

Means, standard deviations, and correlations are presented in Table 5.

**Feedback Goal Priority.** Contrary to our predictions, an independent samples t-test revealed that participants showed no significant difference in how much they prioritized kindness goals in feedback to Sarah than to Andrew, *t*(525) = .22, *p* = .83, *d* = 0.02, 95% *CI* [-0.15, 0.19] (see Figure 1). Also unexpectedly, people rated candor as significantly more of a priority for Andrew than for Sarah, *t*(525) = 2.34, *p* = .02, *d* = 0.20, 95% *CI* [0.03, 0.38]. Again, we note that this study was conducted during the COVID-19 pandemic, which might have shifted the priority of kindness goals when giving feedback to poor performers, regardless of gender, which could explain the difference in the pattern of findings here compared to the earlier studies. Alternatively, given our conception of candor as a potential indicator of hostile sexism which could suggest this gender ideology came into play (but see SOM, there is no moderation by participant gender).

**Mechanisms.** As predicted, we found a significant difference on warmth. Participants rated Sarah significantly higher on warmth than Andrew, *t*(525) = -3.0, *p* = .003, *d* = -0.26, 95% *CI* [-0.43, -0.09]. No gender condition differences emerged on any of the alternate mechanisms: competence (*p* = .35), expectancy violation (*p* = .29), discomfort (*p* = .37), fear of conflict (*p* = .51), perceived emotions (*p* = .10), benevolent sexism (*p* = .38), IMS (*p* = .14), and EMS (*p* = .26).

**Indirect Effect on Kindness Goal Priority.** While the predicted direct effect did not emerge on the goal priority of kindness, this does not preclude the possibility of an indirect effect from gender condition to kindness via our hypothesized mechanism, warmth. Indeed, warmth significantly correlated with kindness ratings, *r* = .36, *p* < .001. To test this, and to evaluate warmth against the alternative mechanisms, we used Process Model 4, 10,000 bootstrap samples, with condition as X, kindness goal priority as Y, and as M all of the possible mechanisms: warmth, competence, expectancy violation, discomfort, fear of conflict, perceived emotions, benevolent sexism, IMS, and EMS. The only indirect effect supported was the hypothesized one, all other 95% CIs crossed zero. As predicted, gender condition, reading about Sarah rather than Andrew, significantly increased perceptions of warmth which was, in turn, associated with higher ratings of kindness as a goal priority, *effect* = .03, bootstrap *SE* = .01, 95% *CI* [.01, .06].

**Indirect Effect on Candid Goal Priority.** We also explored whether any of the mechanisms measured would explain the unexpected effect that emerged on candor, using the same model as above but with candor as Y. No indirect effects were supported.

**Study 4b**

**Method**

**Participants**

This study was conducted via the behavioral lab at a UK Business School and pre-registered (blinded for review). The target sample size was 600 participants from the online community sample in the United Kingdom, though as this was a convenience sample the sample size was determined by how many participants responded. All participants who failed one or more of the a priori exclusion criteria (failed the attention checks, failed to accurately recall the employee name), were excluded from the analysis, leaving 463 participants (315 ‘woman,’ 147 ‘man,’ 1 ‘trans;’; mean age = 34.32, SD = 12.99; 24 ‘Black/African,’ 15 ‘Latino/Hispanic,’ 83 ‘East Asian,’ 1 ‘Native,’ 13 ‘South East Asian,’ 80 ‘South Asian,’ 10 ‘Middle Eastern,’ 225 ‘White/Caucasian,’ 25 ‘Other,’ note that participants could choose multiple options for both gender and race identifications).

**Procedure.** Participants read a similar scenario as in Study 4a: they were about to give feedback to a poor performer [Andrew/Sarah] who had scored a 25% competency level. However, given with our intuition that assumptions about how the pandemic has affected business might shape people’s responses to a poorly performing employee, in this version of the scenario we specified that, “Given that people are on their electronic devices more than ever these days, the past year has been a boom time for selling ad space. Business has been better than ever, despite the challenges of the past year.” We also stated, “Today you have to review each employee’s performance rating and give them evaluative feedback face-to-face in their performance review over zoom,” to evoke a situation more similar to the earlier studies. Participants completed the same measures of feedback goal priority (kindness alpha=.76, candor alpha=.64), warmth (alpha=.91), competence (alpha=.93), discomfort (r=.66, p<.001 ), and fear of conflict as in Study 4a.

**Results**

Means, standard deviations, and correlations are presented in Table 6.

**Feedback Goal Priority.** An independent samples t-test replicated the predicted effect again. Participants indicated that they would prioritize kindness goals more during feedback to Sarah than to Andrew, *t*(463) = -2.14, *p* = .03, *d* = -0.20, 95% CI [-0.38, -0.02] (see Figure 1). Also consistent with earlier studies, people did not prioritize candid goals significantly more for Sarah than Andrew, *t*(463) = 1.26, *p* = .21.

**Mechanisms.** As predicted, we found a significant difference on warmth. Participants rated Sarah significantly higher on warmth than Andrew, *t*(462) = -2.10, *p* = .04, *d* = -0.20, 95% *CI* [-0.38, -0.01]. No gender condition differences emerged significantly on competence (*p* = .06), though there was a trend toward Sarah being rated higher than Andrew, or on discomfort (*p* = .85), or fear of conflict (*p* = .57).

**Indirect Effect on Kindness Goal Priority.** Again, warmth significantly correlated with kindness ratings, *r* = .38, *p* < .001. We used Process Model 4, 10,000 bootstrap samples, with condition as X, kindness goal priority as Y, and as M all of the possible mechanisms measured in this study: warmth, competence, discomfort, and fear of conflict. The only indirect effect supported was the hypothesized one, all other 95% CIs crossed zero. As predicted, gender condition, reading about Sarah rather than Andrew, significantly increased perceptions of warmth which was, in turn, associated with higher ratings of kindness as a goal priority, *effect* = .02, bootstrap *SE* = .01, 95% *CI* [.002, .05].

**Indirect Effect on Candid Goal Priority.** No indirect effects of gender condition via the possible mechanisms on candid goal priority were supported (all *CIs* crossed 0).

**Study 4 Summary**

We found support for the proposed theoretical model in two confirmatory pre-registered studies. Feedback givers who read their poorly performing employee was a woman, rather than a man, saw her as warmer which was associated with placing a higher priority on kindness as a feedback goal. Moreover, this model was supported accounting for alternative mechanisms, ranging from competence perceptions (e.g., if people were overtly sexist), to expectancy violation (e.g., a shifting standards account), to discomfort with giving feedback (e.g., if kindness arose from worry about seeming biased against women), to fear of conflict (e.g., if people withdraw kindness from men who are stereotyped as disagreeable), to perceived emotionality (e.g., which might evoke more kindness), benevolent sexism (e.g., sexist protective instincts), IMS, and EMS (e.g., again relating to the worry about seeming biased). None of these potential alternatives was supported.

While the current studies documented our hypothesized process by measuring warmth directly, our next study was more indirect. Our theory follows the Stereotype Content Model, predicting that people prioritize kindness more as a function of gender stereotypes about warmth. Following this logic, people may use the feedback style that they believe would help the recipient most. To test this, our final study asked how helpful real-world managers see kind feedback as being. Consistent with the earlier studies, we predicted that people would rate kind feedback as more helpful for women than men. We also again tested for evidence of evaluators positively biasing the style of their feedback.

**Study 5**

**Method**

**Participants**

An a priori sample size calculation indicated that a total sample size of N = 190 would provide sufficient power (based on effect size d = .41 from the focal t-test in Study 3, which at the time of the study was appropriate to use, two-tailed, alpha error probability = .05, power = .8, and n2/n1 = 1). We collected a panel sample of 201 participants who were currently in managerial or supervisory roles in the United States and the United Kingdom on Prolific. All participants who failed one or more of the a priori exclusion criteria (failed the attention check, failed to accurately recall the employee name, or through an experimenter error were allowed to take over 30 minutes to complete the survey and thus did not complete it in one sitting), were excluded from the analysis, leaving 156 participants (80 ‘female’, 1 ‘other’; *Mage* = 38.4, *SD* = 10.52; 143 identified as White, 9 as Asian or Asian British, 4 as Black, Caribbean, African, or Black British, and 6 as Mixed or Other; note participants could choose multiple options).

**Procedure.** Participants read the same scenario as in Study 3: they were about to give feedback to a poor performer [Andrew/Sarah] who had scored a 25% competency level.

**Manipulation Check.** Participants were again asked to provide the name of the employee they had just read about.

**Feedback Helpfulness.** We adapted the measure used in Studies 2-3 to investigate participants’ perceptions of the helpfulness of kindness versus candor in improving performance. Participants read: “When giving feedback to Andrew/Sarah, different approaches may be more or less helpful for their future performance.  How helpful do you believe each of the following would be for [Andrew’s/Sarah’s] future performance?” The kindness measure asked participants how helpful to the poor performer’s future performance it would be ‘To be kind and considerate of [Andrew’s/Sarah’s] feelings’, ‘to have a sympathetic and compassionate approach in your feedback to [Andrew/Sarah]’, to ‘focus primarily on [Andrew’s/Sarah’s] performance strengths’, and to ‘focus on increasing [Andrew’s/Sarah’s] confidence and morale’, (alpha = .66). The candid measure asked participants how helpful it would be to the poor performer’s future performance ‘to be completely straightforward and blunt to [Andrew/Sarah] about poor performance’, ‘to be direct and straight to the point in delivery of the feedback to [Andrew/Sarah]’, to ‘focus primarily on [Andrew’s/Sarah’s] performance weaknesses’, and to ‘remind [Andrew/Sarah] of the consequences of poor performance’, (scale: 1 “*Not at all*” - 6 “*Extremely,”* alpha = .61). Item order was randomized.

**Intended feedback.** Participants were asked “what percentage of your feedback do you think should be kind or critical to [Andrew/Sarah]?” They were asked to indicate the proportion of both kind and critical feedback (adding up to 100%) they would give the employee during feedback.

Finally, participants completed additional exploratory measures (see SOM), a standard demographics form, and were debriefed.

**Results**

Means, standard deviations, and the correlations are presented in Table 7.

**Feedback Helpfulness.** As predicted, participants rated kindness as more helpful when randomly assigned to anticipate giving feedback to Sarah than to Andrew, *t*(154) = -2.39, *p* = .02, *d* = -0.38, 95% CI [-0.70, -0.07] (see Figure 1). As in previous studies, participants rated candor as equally helpful for Sarah and Andrew, *t*(154) = .38, *p* = .71, *d* = 0.06, CI [-.25, .37], again providing an indirect indication that hostile sexism is not driving these effects.

**Intended feedback.** Participants’ ratings of how helpful kind feedback is correlated significantly with the proportion of their intended feedback they anticipated to be kind (vs. critical), *r* = .51, *p* < .001. We analyzed the data using Model 4 in Process (Hayes, 2012), with 20,000 bootstrapped samples, entering Gender Condition as X, Intended Feedback (proportion of positive vs critical content) as Y, and Helpfulness of Kind Feedback as M. The indirect effect was supported: Anticipating giving feedback to a woman (vs. man) increased participants’ perception that kind feedback is helpful, which increased their likelihood of indicating they would give a higher proportion of positive, rather than critical, feedback, effect = 3.16, bootstrap *SE* = 1.35, 95% CI [.57, 5.87].

**Mini Meta-Analysis**

We have reported the results of five experimental studies that test the same directional hypothesis: that people would prioritize kindness more as a goal when anticipating giving developmental feedback to a woman (vs. man). Our file drawer consists of 2 additional studies, one using Amazon mTurk (Study SOM1) and one using a lab community sample (Study SOM2), to test the same directional hypothesis using the same manipulation, method, and dependent variable as in Study 3. Study SOM1 found a nonsignificant trend toward higher priority of kindness goals for Sarah than for Andrew, *t*(387) = 1.54, *p* = .12, *d* = .16, 95% CI [-.04, .34]. Study SOM2 found a nonsignificant trend toward higher priority of kindness goals for Andrew than for Sarah, *t*(322.79) = 1.61, *p* = .11, *d* = .18, 95% CI [-.04, .39]. In cases such as these, where we have 6 experimental studies, with 3 reporting a significant effect (with varying effect sizes, Studies 2, 3, and 4b) and 3 that found nonsignificant effects (Studies 4a, though an indirect effect emerged via warmth, SOM1, which found a trend in the predicted direction, and SOM2, which found a trend in the opposite direction) while testing the same directional hypothesis, a mini meta-analysis (Goh et al., 2016) offers the ability to evaluate the direction and reliability of the effect and, if one exists, to more accurately estimate its size. We followed the procedures outlined in Goh et al. (2016) and used the associated annotated (updated) spreadsheet to estimate the mean effect size using fixed effects and inverse variance weighting. We entered the Cohen’s *d* for each study and sample sizes by condition: Study 2 *d* = .26, N Sarah = 176, N Andrew = 173; Study 3 *d* = .40, N Sarah = 48, N Andrew = 46; Study 4a *d* = .02, N Sarah = 217, N Andrew = 256; Study 4b *d* = -.20, N Sarah = 235, N Andrew = 230, Study 5 *d* = -.38, N Sarah = 79, N Andrew = 77; Study SOM1 *d* = -.17, N Sarah = 191, N Andrew = 177; and Study SOM2 *d* = .18, N Sarah = 170, N Andrew = 162. Across the 6 studies, gender reliably influenced the prioritization of kindness as a feedback goal in the hypothesized direction, mini meta-analysis effect size *d* = -.12, se = .04, *z* = -2.83, *p* = .004, 95% CI [-.20, -.04], two-tailed. Thus, the results of this mini meta-analysis offer support in favor of the directional hypothesis and estimate a small effect size, by which people are biased toward prioritizing kindness as a feedback goal when giving developmental performance feedback to a woman (vs. man).

**General Discussion**

Across 6 studies using diverse samples (professionals, MBA students, lab participants from the community, and managers), we find that people place greater priority on the goal of being kind in developmental feedback to women than to men. Study 1 showcased this pattern among professionals (supervisors, peers, and subordinates) from around the world giving developmental feedback to their highly performing former colleagues, now enrolled in an MBA, on how they could improve. Study 2 found that a highly international group of MBA students prioritize kindness more as a feedback goal when anticipating giving developmental feedback to a woman (vs. man) employee who had performed relatively poorly, but “above bar.” Study 3 replicated this result among a lab sample even after standardizing the level of performance as low. Studies 4a and 4b investigated mechanism and found that a panel of individuals with supervisory experience (4a) and community sample (4b) who were anticipating giving feedback to a poorly performing woman, rather than a man, perceived her as warmer and thus increased the priority they placed on kindness as a feedback goal. These studies also ruled out the alternate potential mechanisms of competence, shifting standards through expectancy violation, discomfort with giving feedback, fear of conflict, perceived emotions, benevolent sexism, and internal and external motivation to control prejudice. Study 5 conceptually replicated the core pattern, showing that managers view kindness in developmental feedback as more helpful to improving a woman’s poor performance than a man’s. A mini-meta analysis across the experimental studies and two file-drawer studies offered further support of our primary hypothesis and a better estimate of the effect size, which could be evaluated as small. Thus, regardless of whether they have just given feedback to high-achieving MBAs (Study 1) or are anticipating giving feedback to a hypothetical employee with either adequate (Study 2) or low performance (Studies 3-5), evaluators exhibited a bias toward prioritizing the goal of kindness to women versus men.

We theorized that this kindness bias helps explain positivity biases in feedback to women versus men. In Study 1, evaluators giving developmental feedback to women (rather than men) rated kindness as a higher feedback goal priority which was in turn associated with the positive emotion words included in their qualitative feedback to MBAs. Experimentally, indirect effects consistent with positively inflated feedback emerged in Studies 3 and 4. Although these indirect effects across studies offer a consistent pattern, we must note that the confidence intervals are consistently close to zero. We also found that people do not reliably prioritize candor more (Studies 1-3, 4b), though this effect emerged in Study 4a, or see candor as more helpful (Study 5) for women than men. This implies that gendered differences exist in kindness goals despite people’s generally equal prioritization of honesty as a goal in developmental feedback.

**Theoretical Implications**

Our work advances the study of gender bias in the workplace by identifying goals as part of the psychological process that shape gendered outcomes in the workplace, extending the Stereotype Content Model. Future research may benefit from more precisely identifying the cognitive processes by which warmth stereotypes drive these effects – whether through the cognitive overlap of salient constructs (the gender stereotype of warmth and the feedback goal of kindness) as we theorize, or perhaps through more complex cognitive processes, such as anticipatory reciprocity whereby people match expectations of another’s behavior. More broadly, our findings indicate that focusing on evaluators' goals, in addition to their implicit biases (the topic of previous interventions; e.g., Devine et al., 2012) and their stereotypes of women’s competence (Kray et al., 2014), may be necessary to reduce gender bias in essential and everyday workplace processes. Future research should explore whether goal-setting interventions can mitigate bias in people’s goal priorities and, subsequently, their performance feedback. These interventions could further evaluate whether the processes we identify here vary based on the performance of the feedback recipient. While our studies observed the proposed effects among both high (Study 1) and low (Studies 2-5) performers, direct comparisons across performance levels in organizational contexts where performance reputations build over time might reveal nuances. We find it notable that, in contrast to past work (e.g., King et al., 2012), our findings consistently emerge among both women and men. This suggests that any interventions developed by future research to address the gendered goal priorities identified in this work must be equally effective across the spectrum of gender.

These findings also raise important questions about who loses—and how—as a consequence of people prioritizing kindness more as a goal when giving developmental feedback to women versus men, especially if candor is equally prioritized. If greater kindness to women shrouds the candor of the feedback through positive wording, it may inhibit women’s ability to learn and predict their future outcomes, which women in devalued positions report as problematic (Vescio et al., 2005). However, less kindness in feedback to men may foster cultures that reduce men’s dedication and well-being at work (e.g., Glick et al., 2018). Future research should investigate how gendered kindness differentials are experienced, whether they obscure candor, and how they may contribute to fostering cultures of toxic masculinity that are constraining to all genders (Glick et al., 2015; Ringov & Zollo, 2007; Rudman & Glick, 1999). For example, research in a field setting could test whether manipulating feedback-givers’ goals (e.g., by encouraging them to be kind in their feedback) elicits more equal feedback across genders, and whether it influences feedback recipients’ performance over time differentially (or not) by gender.

**Limitations and Future Directions**

Our work used names that likely activated a European or White prototype. Future research must address questions about how intersections of race and gender shape feedback goal priorities (e.g., Rattan et al., 2017) and also explore intersectional and gender-nonbinary dynamics. For example, goals of both kindness and not appearing racist (Harber, 1998) may come to be activated when evaluators give feedback to racially minoritized women, which could exacerbate biases. Research considering peoples’ responses to trans and gender-nonbinary feedback recipients would also shed further understanding, and perhaps complexity, on the dynamics investigated here. It would also be interesting for future research to examine whether women’s emotion displays in response to feedback (e.g., counterstereotypical anger, or stereotypical hurt, vice versa for men) moderate the degree to which people exhibit the kindness bias, as well as whether such emotion displays might bring gender ideologies such as benevolent or hostile sexism back into play. In addition, research should explore whether people who exhibit a meta-knowledge of obstacles for women in the workplace (e.g., confidence gaps, Kling et al., 1999; opportunity gaps, Heilman, 2001; belonging gaps, Good et al., 2012) might ironically be even more prone to the patterns we document in this research because of a desire to be inclusive toward, or support the workplace belonging of, women. We must also note that the moderate reliability of our measures of kindness and candor suggest further improvements to our measures could be made. We encourage future research to do so in the context of actual feedback, which would allow for tracking the real-life consequences of these feedback dynamics.

**Conclusion**

Whether people’s goals are for developmental feedback to be more or less kind, it ought to be as equal as possible across gender. Our results suggest that it may be necessary to either increase how much people prioritize the feedback goal of kindness toward men, or keep it in check toward women, in order to achieve the broader social goal of gender equity in workplace feedback. Thus, the present research suggests that taking into account the goal priorities of the person giving the feedback, and how these feedback goal priorities are shaped by the recipient’s gender, may be essential for fully understanding, and addressing, the sources of gender biases in the workplace.

**References**

Annett, J. (1969). Feedback and human behaviour: The effects of knowledge of results, incentives and reinforcement on learning and performance. Penguin Books. Retrieved from http://psycnet.apa.org/psycinfo/1970-19815-000.

Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of personality and social psychology,* 95(4), 918.

Barney, J. B. (1986). Organizational culture: Can it be a source of sustained competitive advantage? *Academy of Management Review, 11*(3), 656–665. http://doi.org/10.5465/AMR.1986.4306261.

Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology, 102*(6), 1214–1238. http://doi.org/10.1037/a0027717

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66*(1), 5–20. http://doi.org/10.1037/0022-3514.66.1.5.

Biernat, M., Tocci, M. J., & Williams, J. C. (2012). The language of performance evaluations: Gender-based shifts in content and consistency of judgment. *Social Psychological and Personality Science, 3*(2), 186–192. http://doi.org/10.1177/1948550611415693.

Correll, S. J., Weisshaar, K. R., Wynn, A. T., & Wehner, J. D. (2020). Inside the black box of organizational life: The gendered language of performance assessment. *American Sociological Review, 85*(6), 1022-1050.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of personality and social psychology, 92*(4), 631.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology 40*, 61-149. <http://doi.org/10.1016/S0065-2601(07)00002-0>.

Eagly, A. H., & Wood, W. (1982). Inferred sex differences in status as a determinant of gender stereotypes about social influence. *Journal of Personality and Social Psychology, 43*(5), 915–928. http://doi.org/10.1037/0022-3514.43.5.915.

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.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275–298. https://doi.org/https://doi.org/10.1146/annurev-psych-122216- 011719.

Fisher, C. D. (1979). Transmission of positive and negative feedback to subordinates: A laboratory investigation*. Journal of Applied Psychology, 64*(5), 533–540. http://doi.org/10.1037/0021-9010.64.5.533.

Fisher, C. D. (1984). Transmission of positive and negative feedback to subordinates: a lab investigation. *Journal of Applied Psychology, 64*(5), 533.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. http://doi.org/10.1037/0022-3514.82.6.878.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491–512. http://doi.org/10.1037/0022-3514.70.3.491.

Glick, P., Berdahl, J. L., & Alonso, N. M. (2018). Development and validation of the masculinity contest culture scale. *Journal of Social Issues, 74*(3), 449-476.

Glick, P., Wilkerson, M., & Cuffe, M. (2015). Masculine identity, ambivalent sexism, and attitudes toward gender subtypes: Favoring masculine men and feminine women. *Social Psychology, 46*(4), 210–217. http://doi.org/10.1027/1864-9335/a000228.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta‐analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10*(10), 535-549.

Good, C., Rattan, A., & Dweck, C. (2012). Why do women opt out? Sense of belonging and women’s representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700. Retrieved from http://psycnet.apa.org/journals/psp /102/4/700/.

Harber, K. D. (1998). Feedback to minorities: evidence of a positive bias. *Journal of Personality and Social Psychology, 74*(3), 622–628. http://doi.org/10.1037/0022-3514.74.3.622.

Harber, K. D., Stafford, R., & Kennedy, K. (2010). The positive feedback bias as a response to self-image threat. *The British Journal of Social Psychology /The British Psychological Society, 49*(1), 207–218. http://doi.org/10.1348/014466609X473956

Harris, M. M., & Schaubroeck, J. (1988). A meta‐analysis of self‐supervisor, self‐peer, and peer‐supervisor ratings. *Personnel Psychology, 41*(1), 43-62.

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. http://doi.org/10.1111/0022-4537.00234.

Heisig, J. P., & Schaeffer, M. (2019). Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *European Sociological Review*, *35*(2), 258-279.

Jampol, L ., & Zayas, Z. (2020). Gendered white lies: Performance feedback is upwardly distorted to women. Unpublished manuscript.

King, E. B., Botsford, W., Hebl, M. R., Kazama, S., Dawson, J. F., & Perkins, A. (2012). Benevolent sexism at work. *Journal of Management, 38*(6), 1835–1866. http://doi.org/10.1177/0149206310365902.

Kling, K., Hyde, J., & Showers, C. (1999). Gender differences in self-esteem: a meta-analysis. *Psychological Bulletin, 125*(4), 470. Retrieved from http://psycnet.apa.org/psycinfo/1999-05876-006.

Kray, L. J., Kennedy, J. A., & Van Zant, A. B. (2014). Not competent enough to know the difference? Gender stereotypes about women’s ease of being misled predict negotiator deception. *Organizational Behavior and Human Decision Processes, 125*(2), 61–72. http://doi.org/10.1016/j.obhdp.2014.06.002

Lean in, & McKinsey & Company. (2015). Women in the workplace. Leanin.org. Retrieved from http://womenintheworkplace.com/ui/pdfs/Women\_in\_the\_Workplace\_2015.pdf?v=5

Locke, E. A., Latham, G. P., & Edwin, A. (2002). Building a practically useful theory of goal setting and task motivation, *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037//0003-066X.57.9.705.

London, M. (2003). Job feedback: Giving, seeking, and using feedback for performance improvement. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=xhKDnSIGdlYC&oi=fnd&pg=PT12&dq=managers+and+giving+performance+feedback&ots=oPdW7fe4gZ&sig=lGvGeoZp76zpypdDnxYcttXD6DQ.

Lupoli, M. J., Jampol, L., & Oveis, C. (2017). Lying because we care: compassion increases prosocial lying because we care: compassion increases prosocial lying. *Journal of Experimental Psychology: General, 146*(7), 1026. http://doi.org/10.1037/xge0000315.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty’s subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474–16479. http://doi.org/10.1073/pnas.1211286109.

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates. www.LIWC.net.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, *40*(3), 879-891.

Rabe-Hesketh, S., & Skrondal, A. (2012). Multilevel and longitudinal modeling using Stata-volume II: Categorical responses. STATA Press.

Rattan, A., Steele, J., & Ambady, N. (2019). Identical applicant but different outcomes: The impact of gender versus race salience in hiring. *Group Processes & Intergroup Relations, 22*(1), 80-97.

Ringov, D., & Zollo, M. (2007). The impact of national culture on corporate social performance. *Corporate Governance, 7*(4), 476-485. Retrieved from http://www.emeraldinsight.com/doi/pdf/10.1108/14720700710820551.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology, 77*(5), 1004–1010. http://doi.org/10.1037/0022-3514.77.5.1004.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology, 48(*1), 165-179.

Sørensen, J. B., & Sorensen, J. B. (2002). The strength of corporate culture and the reliability of firm performance. *Administrative Science Quarterly, 47*(1), 70. http://doi.org/10.2307/3094891.

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin, 105*(3), 409–429. http://doi.org/10.1037/0033-2909.105.3.409.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*.

Tesser, A., & Rosen, S. (1975). The reluctance to transmit bad news. *Advances in Experimental Social Psychology, 8*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0065260108602518.

Vescio, T. K., Gervais, S. J., Snyder, M., & Hoover, A. (2005). Power and the creation of patronizing environments: The stereotype-based behaviors of the powerful and their effects on female performance in masculine domains. *Journal of Personality and Social Psychology, 88*(4), 658–672. http://doi.org/10.1037/0022-3514.88.4.658.

Waung, M., & Highhouse, S. (1997). Fear of conflict and empathic buffering: Two explanations for the inflation of performance feedback. *Organizational Behavior and Human Decision Processes, 71*(1), 37-54.

Wiltermuth, S. S., Newman, D. T., & Raj, M. (2015). The consequences of dishonesty. *Current Opinion in Psychology, 6*, 20–24. http://doi.org/10.1016/j.copsyc.2015.03.016.

**Tables and Figures**

**Table 1.** Study 1, Raters’ Self-Reports of Priority for Kindness Goals and Candor Goals.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | DV: Rater Priority of Kindness | | DV: Rater Priority of Kindness | |  | DV: Rater Priority of Candor | | DV: Rater Priority of Candor | |
| Variable | Estimate | *SE* | Estimate | *SE* | | Estimate | *SE* | Estimate | *SE* |
| Intercept | 3.10 | 0.03 | 2.46 | 0.31 | | 4.36 | 0.02 | 4.22 | 0.16 |
| Recipient Gender | 0.09† | 0.05 | 0.11\* | 0.05 | | -0.05† | 0.03 | -0.04† | 0.03 |
| Peer Relationship |  |  | 0.13\*\* | 0.05 | |  |  | -0.05 | 0.03 |
| Direct Report Relationship |  |  | 0.23\*\*\* | 0.06 | |  |  | 0.05 | 0.04 |
| Customer Relationship |  |  | 0.03 | 0.06 | |  |  | -0.03 | 0.04 |
| Recipient Age |  |  | 0.02† | 0.01 | |  |  | 0.006 | 0.005 |
|  |  |  |  |  | |  |  |  |  |
| *n* | 4,371 | | 4,341 | | | 4,371 | | 4,341 | |
| *ψ* | 0.01 | | .10 | | | 0.01 | | .01 | |

*Note:* † *p*<.10 *\*p*<.05 *\*\*p*<.01 *\*\*\*p*<.001*.* Gender was coded with men as the reference group. Relationship was coded with supervisor ratings as the reference group. ψ represents the random intercept variance.

**Table 2.** Study 1,Qualitative Analysis of Feedback Style as a function of Recipient Gender.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | DV: Tone of Qualitative Feedback | | | DV: Positive Emotions in Qualitative Feedback | | DV: Negative Emotions in Qualitative Feedback | |
| Variable | | Estimate | *SE* | Estimate | *SE* | Estimate | *SE* |
| Intercept | | 97.55 | 7.04 | 6.00 | 1.12 | 0.90 | 0.53 |
| Recipient Gender | | 3.32\*\* | 1.15 | 0.54\*\* | 0.18 | -0.20\* | 0.09 |
| Peer Relationship | | -0.47 | 1.28 | 0.43\* | 0.21 | 0.22\* | 0.10 |
| Direct Report Relationship | | 0.12 | 1.62 | 0.74\*\* | 0.27 | -0.07 | 0.13 |
| Customer Relationship | | -2.12 | 1.50 | 0.21 | 0.25 | -0.004 | 0.12 |
| Recipient Age | | -0.64\*\* | 0.24 | 0.005 | 0.04 | 0.02 | 0.02 |
|  | |  | |  | |  | |
| *n* | | 3,381 | | 3,381 | | 3,381 | |
| *ψ* | | 25.11 | | 0.47 | | 0.07 | |

*Note:* † *p*<.10 *\*p*<.05 *\*\*p*<.01 *\*\*\*p*<.001*.* Gender was coded with men as the reference group. Relationship was coded with supervisor ratings as the reference group. Feedback tone ranged from 0 to 100, with higher values indicating more positivity. Other variables represent the percent of total words in the feedback that were related to that type of emotion. ψ represents the random intercept variance.

**Table 3.** Study 2 means, standard deviations, and correlations.

|  |  |  |  |
| --- | --- | --- | --- |
| Gender Condition | Woman (Sarah, coded as 1, N = 176) | Man (Andrew, coded as 0, N = 173) | r |
| 1. Kind Feedback | 4.08a (0.82) | 3.86b (0.81) |  |
| 2. Candid Feedback | 4.26a (0.88) | 4.37a (0.79) | -0.06 |

*Note*: Across rows, means with different subscripts differ at the *p* ≤ 0.05 level. An asterisk indicates a significant correlation at the *p* ≤ 0.05 level.

**Table 4.** Study 3 means, standard deviations, and correlations.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gender Condition | Woman (Sarah, coded as 1, N = 48) | Man (Andrew, coded as 0, N = 46) | 1 | 2 | 3 | 4 |
| 1. Kind Feedback | 4.42a (0.97) | 4.01b (1.04) |  |  |  |  |
| 2. Candid Feedback | 4.60 (0.85) | 4.45a (0.74) | -0.02 |  |  |  |
| 3. Feedback Valence | 2.94 (1.04) | 2.98 (1.16) | 0.49\* | 0.06 |  |  |
| 4. Feedback Straightforwardness | 3.69 (.90) | 3.89 (1.18) | -.38\* | 0.34\* | -0.37\* |  |
| (rc) |  |  |
| 5. Feedback Focus | 2.38 (1.02) | 2.37 (1.02) | 0.23\* | -0.34\* | 0.22\* | -0.27\* |

*Note*: Across rows, means with different subscripts differ at the *p* ≤ 0.05 level. An asterisk indicates a significant correlation at the *p* ≤ 0.05 level.

**Table 5.** Study 4a means, standard deviations, and correlations.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  | |
| Gender Condition | Woman (Sarah, coded as 1, N = 271) | Man (Andrew, coded as 0, N = 256) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|  |
|  |
| 1. Kind Feedback | 4.31 (0.81) | 4.33 (0.82) |  |  |  |  |  |  |  |  |  |  | |  |
| 2. Candid Feedback | 4.38 (0.70) | 4.54 (0.73) | 0.00 |  |  |  |  |  |  |  |  |  | |  |
| 3. Warmth | 3.41 (0.89) | 3.19 (0.84) | 0.36\*\* | 0.02 |  |  |  |  |  |  |  |  | |  |
| 4. Competence | 2.57 (0.82) | 2.50 (0.87) | 0.28\*\* | -0.01 | 0.62\*\* |  |  |  |  |  |  |  | |  |
| 5.Expectancy Violation | 3.93 (0.84) | 3.85 (0.97) | 0.01 | 0.11\*\* | 0.08 | 0.11\*\* |  |  |  |  |  |  | |  |
| 6. Discomfort | 3.88 (1.33) | 3.98 (1.21) | -0.07 | 0.21\*\* | 0.14\*\* | 0.08 | 0.00 |  |  |  |  |  | |  |
| 7. Fear of Conflict | 3.59 (1.25) | 3.52 (1.30) | 0.16\*\* | -0.11\*\* | 0.05 | 0.06 | 0.04 | -0.41\*\* |  | |  |  | |  |
| 7. Perceived emotions | 2.99 (0.71) | 2.89 (0.73) | 0.05 | 0.12\*\* | 0.00 | -0.10\* | 0.17\*\* | -0.27\*\* | .23\*\* |  |  |  | |  |
| 8. Benevolent Sexism | 2.74 (0.91) | 2.67 (0.95) | -0.05 | 0.09\* | 0.09\* | . 0.09\* | 0.03 | 0.09\* | -0.06 | 0.03 |  |  | |  |
| 9. IMS | 5.21 (0.85) | 5.09 (0.94) | 0.04 | 0.00 | 0.05 | -0.07 | 0.14\*\* | 0.00 | 0.02 | 0.09\* | -0.31\*\* |  |
| 10. EMS | 3.37 (1.19) | 3.25 (1.15) | 0.00 | 0.01 | 0.01 | -0.03 | 0.03 | -0.05 | 0.05 | .11\* | .25\*\* | 0.04 | |  |

*Note:* Across rows, means with different subscripts differ at the *p* ≤ 0.05. An asterisk indicates a significant correlation at the *p* ≤ 0.05 level and two asterisks indicate a significant correlation at the p ≤ 0.01 level.

**Table 6.** Study 4b means, standard deviations, and correlations.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| Gender Condition | Woman (Sarah, coded as 1, N = 235) | Man (Andrew, coded as 0, N = 230) | 1 | 2 | 3 | 4 |
|  |
|  |
| 1. Kind Feedback | 4.41 (0.80) | 4.24 (0.90) |  |  |  |  |  |
| 2. Candid Feedback | 4.34 (0.77) | 4.43 (0.72) | 0.03 |  |  |  |  |
| 3. Warmth | 3.14 (0.93) | 2.96 (0.92) | 0.38\*\* | -0.01 |  |  |  |
| 4. Competence | 2.58 (0.92) | 2.43 (0.87) | 0.35\*\* | 0.01 | 0.65\*\* |  |  |
| 5. Discomfort | 3.54 (1.30) | 3.51 (1.35) | -0.03 | .24\*\* | 0.12\*\* | .15\*\* |  |

*Note:* Across rows, means with different subscripts differ at the *p* ≤ 0.05. An asterisk indicates a significant correlation at the *p* ≤ 0.05 level and two asterisks indicate a significant correlation at the p ≤ 0.01 level.

**Table 7.** Study 5 means, standard deviations, and correlations.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Factor | Woman (Sarah, coded as 1, N = 79) | Man (Andrew, coded as 0, N = 77) | 1 | 2 |
| 1. Kind Feedback | 4.73 a (0.80) | 4.43b (0.80) |  |  |
| 2. Candid Feedback | 3.87a (0.93) | 3.93 (0.90) | -0.11 |  |
| 3. Intended Feedback | 46.43 (16.85) | 43.26 (15.89) | 0.36\* | -0.16\* |

*Note*: Across rows, means with different subscripts differ at the *p* ≤ 0.05 level. An asterisk indicates a significant correlation at the *p* ≤ 0.05 level.

**Figure 1.**

Chart, bar chart

Description automatically generated

*Note*: Results for Studies 2, 3, 4a, 4b, and 5. The effect of employee gender condition on feedback goals (kind or candid) across experiments. Bars represent mean agreement on a 1-6 scale with feedback goal priority of kind and candid feedback approaches. Error bars are ± SEM.

1. Note that not every rater provided answers to every question. [↑](#footnote-ref-1)