https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–20
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Gender and Preferences for Performance Feedback

Katherine Coffman,[a,*] **David Klinowski**[b]

[a] Harvard Business School, Boston, Massachusetts 02163; [b] Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

*Corresponding author

Contact: kcoffman@hbs.edu, https://orcid.org/0000-0003-4293-2220 (KC); dklinowski@katz.pitt.edu (DK)

Received: August 8, 2023
Revised: October 19, 2023; January 5, 2024
Accepted: January 10, 2024
Published Online in Articles in Advance: August 7, 2024

https://doi.org/10.1287/mnsc.2023.02482

Abstract. Across multiple studies, we investigate whether there are gender differences in preferences for receiving performance feedback. We vary many features of the feedback context: whether the performance task is a cognitive test or a mock interview, the feedback is objective or subjective, and it is possible for the provider of the feedback to discriminate on the basis of gender. Consistent with past work, we find that women are less optimistic about their performance than men and that, on average, more optimistic individuals have greater demand for feedback. Results like these have been hypothesized in the literature to imply that women will shy away from performance feedback more so than men. And, when we survey participants from a similar population, they also anticipate that women will demand feedback at lower rates than men. Yet, across our two incentivized studies, we find that women are no less eager to receive performance feedback than men. Understanding whether and how these results might generalize to broader contexts, particularly those with more social factors, is an important question for future work.

History: Accepted by Dorothea Kubler, behavioral economics and decision analysis.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2023.02482.

Keywords: economics: behavior and behavioral decision making • economics: microeconomic behavior • gender

1. Introduction

Performance feedback is prevalent in educational and professional settings. Peers, teachers, bosses, and mentors often have both the knowledge and the opportunity to offer feedback about an individual's strengths and weaknesses. This type of feedback may inform decisions about what educational tracks and career paths to pursue. Of course, in many important contexts, opportunities for feedback are not exogenously assigned: individuals can seek out, or avoid, performance feedback. These decisions about whether, when, and how intensively to pursue performance feedback may have implications for the quantity and quality of information an individual has about one's own talents. Furthermore, preferences over whether to receive or avoid feedback may shape what opportunities and paths an individual pursues in the first place; a feedback-averse individual may choose to avoid careers that entail a lot of feedback, driving sorting into certain educational tracks and careers. For instance, someone who enjoys feedback may opt into more competitive, results-oriented industries or roles. Moreover, the management literature has found that "feedback-seeking behavior" (Ashford et al. 2003) might, in and of itself, improve a person's labor market outcomes, as "[f]eedback-seeking behavior […] has been linked to higher job satisfaction, greater creativity on the job, faster adaptation in a new organization or role, and lower turnover" (Stone and Heen 2014, p. 9). For these reasons, understanding demand for performance feedback may be valuable for understanding economic outcomes.

In this paper, we explore preferences for feedback, focusing on gender differences. Cleanly identifying gender differences in demand for feedback is challenging using observational data, which is potentially plagued by selection and confounds. Instead, we conduct a series of controlled experiments in which participants complete a task, report their beliefs about their performance, and make incentivized decisions about whether to receive feedback on their performance. In addition, we elicit beliefs from a separate sample about whether there are gender differences in demand for feedback, for our tasks and more broadly.

We explore demand for feedback across a range of settings, including examining different tasks, manipulating the difficulty of the task, allowing the possibility of gender-based discrimination by the provider of the feedback, and examining demand for both objective and subjective feedback. In each of these settings, we find that women demand performance feedback

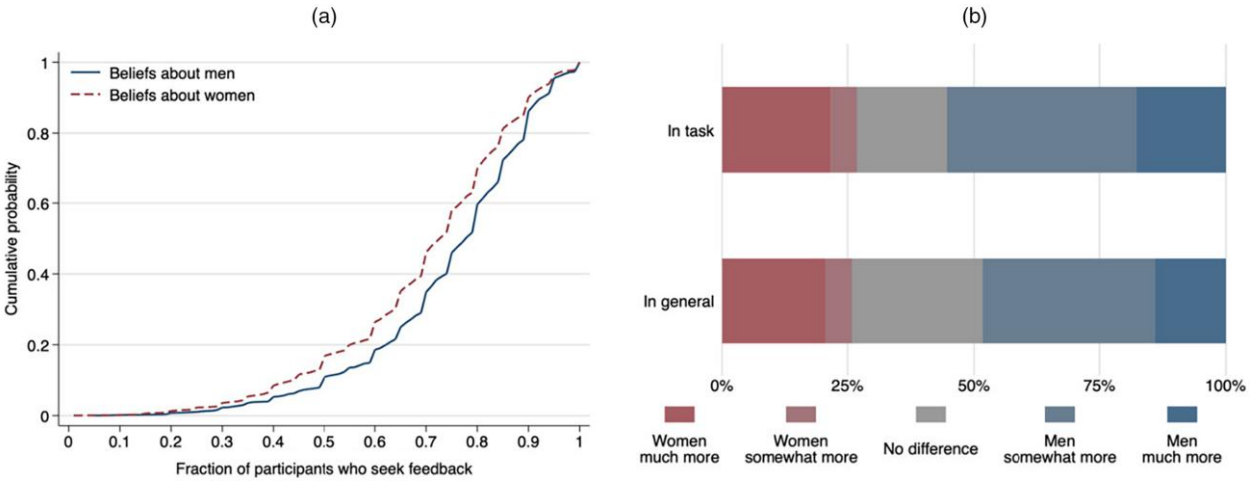
no less than men. This is notable given that our studies incorporate several aspects that ex ante might be expected to lead women to disproportionately shy away from feedback, including receiving feedback on more stereotypically male-typed skills and in domains where they are less self-confident than men. In fact, when we ask different participants from a similar population to predict our results, they believe that women will demand performance feedback less than men. Yet, across a range of regression specifications with and without controls, we consistently find that women's demand for feedback is no less than that of men.

Our experiments share a common format: participants complete a task, and we elicit their demand for performance feedback on that task. In study 1, participants take a cognitive skills test featuring questions on math, general science, and mechanical comprehension, exogenously varying the difficulty of the test across participants. We elicit incentivized demand for learning their absolute and relative performance on the test. In study 2, participants provide written answers to three questions about their own life achievements and personality that are commonly asked in job interviews. We then elicit incentivized demand for learning their relative performance in terms of assessed intellectual curiosity, tendency to strive for achievement, assertiveness, and tolerance to stress. These assessments are made by human resources (HR) professionals who we hire, and the assessments are based on the written answers participants provide. We exogenously vary whether the HR professionals observe the gender of the participants when evaluating their answers. This varies whether discrimination on the basis of gender is possible across treatments, allowing us to test whether the possibility of discrimination by the feedback provider impacts preferences for feedback.

In both studies, we collect individuals' ex ante beliefs about their own performance, including both beliefs of how they performed and how certain they are about those beliefs. Across our two studies, we find that participants who have more optimistic beliefs of their performance, and who are more certain of their beliefs, have greater demand for feedback. Although we see evidence of gender differences in beliefs about own performance (men are more optimistic than women), this does not produce a gender gap in demand for feedback. On net, women are no more feedback averse than men.

We ask individuals from a separate population to predict our study participants' demand for feedback. We collect incentivized beliefs of the share of male and female participants who opt for feedback. Figure 1(a) shows that they do anticipate a gap: pooling across all opportunities to seek out feedback, the mean belief is that 70.2% of women and 74.7% of men would opt for feedback across our two studies (t-test $p < 0.001$), and beliefs about men's demand for feedback stochastically dominate beliefs about women's demand ($p < 0.001$ for both Kolmogorov-Smirnov and Somers' D tests). We also collect more qualitative beliefs of gender differences in willingness to seek feedback using unincentivized Likert-style questions. The belief that men have

Figure 1. (Color online) Beliefs of Others' Demand for Feedback



Notes. (a) Cumulative distribution of incentivized beliefs of what fraction of men and women choose to receive feedback on their performance on the task when there is no cost to receive or avoid feedback. (b) Likert: Responses to who demands feedback more on a Likert scale. Top bar (In task) graphs answers to the question: "Overall, for the task participants completed in the previous study, how would you describe differences in men's and women's preferences for finding out how they performed." Bottom bar (In general) graphs answers to the question: "Thinking more generally—not just for the task the participants completed in the previous study—how would you describe gender differences in preferences for finding out how they performed in educational and professional settings, such as tasks in school and work?" Panels (a) and (b) pool observations from both the cognitive test and interview studies.

a greater demand for feedback is approximately twice as common as the belief that women have a greater demand for feedback, both in terms of our study tasks and in educational and professional settings more generally (Wilcoxon signed-rank test, $p < 0.001$ for both). Thus, whereas we see that women in our studies demand performance feedback no less than men, others believe them to be more feedback averse than men.

Our main contribution is to test the hypothesis that women demand less feedback than men, using two well-powered, preregistered experiments designed for this purpose. Whereas this hypothesis has been discussed in the behavioral economics literature, for example, in Niederle and Vesterlund (2007), it has been largely unexplored, with a few important exceptions. In Eil and Rao (2011) and Möbius et al. (2022), subjects can pay to learn or avoid learning their relative performance on an ego-relevant domain after receiving a series of noisy signals. Both studies find no gender differences in the valuation for the information on average, although Möbius et al. (2022) find, in addition, that women are more likely than men to pay to avoid the information.[1] In Castagnetti and Schmacker (2022) and Sharma and Castagnetti (2023), subjects choose between different information structures that give noisy feedback on their relative performance in an ego-relevant domain. Sharma and Castagnetti (2023) find that women prefer more noisy structures than men, suggesting that women are more likely to avoid feedback, but Castagnetti and Schmacker (2022) find no such difference.

We build on this important work by conducting two well-powered, preregistered studies specifically designed to explore gender differences in demand for feedback. Across our studies, we examine the role of confidence in explaining any gender gap in preferences for feedback. We also examine whether the possibility of gender-based discrimination by the provider of feedback impacts men's and women's demand for feedback differently, a factor that has not been studied previously. We complement our experiments with parallel studies investigating beliefs about gender differences in demand for feedback. Together, our studies bring new evidence to the important question of whether men and women vary in their preferences for receiving feedback, laying the groundwork for future investigations.

Finally, we should emphasize that our studies consider specific settings that intentionally shut down some interesting channels through which preferences for feedback may operate. The feedback provided in our study is not particularly actionable, nor does it have significant strategic value. It is provided within a minimalistic context, stripped of many social aspects that may loom large in other environments, such as relational, reputational, and power dynamics between the feedback provider and feedback recipient. Thus, we cannot rule out gender differences in receiving performance feedback in other settings. What we can say from our evidence is that absent these other factors, it does not seem to be the case that women demand feedback less than men. This finding may be important in interpreting and addressing gender gaps in demand for feedback in other settings, as it suggests there may be ways to alter contexts to reduce differences.

2. Hypotheses

This section presents the hypotheses that guide our analysis. Our primary interest in both studies is to test for gender differences in the demand for performance feedback.

Past literature on ego management and self-confidence provides clues as to why women might demand performance feedback less than men. In our studies, feedback has minimal instrumental value, at least within the study, as feedback is received at the end of the experiment and has no value for decision making within the experiment. There is a large body of evidence that information is valued for reasons other than its instrumental use (for reviews, see Golman et al. 2017 or the discussion of the literature by Masatlioglu et al. 2023). One reason is ego management: if individuals derive utility from holding favorable beliefs about themselves, then information that affects these beliefs can affect utility directly.[2] Indeed, Eil and Rao (2011), Burks et al. (2013), Masatlioglu et al. (2023), and Golman et al. (2022) show that individuals who are more optimistic about their performance on an ego-relevant task are more eager to receive performance feedback, indicating an intrinsic preference for news that is positive for self-image.[3]

Combine this insight with the findings of the literature on gender differences in self-confidence. Across a range of studies, researchers have found that women are more pessimistic than men about their own abilities and performance on tasks conditional on true ability (e.g., Niederle and Vesterlund 2007, Grosse and Reiner 2010, Shurchkov 2012, Coffman 2014, Buser et al. 2014, Bordalo et al. 2019, Klinowski 2019, Exley and Kessler 2022, Möbius et al. 2022, Exley and Nielsen 2024). This evidence comes mostly from experiments in which participants perform male-typed, ego-relevant tasks, such as a cognitive skills test. Taken together, the implication would seem to be that women are more feedback averse than men: if more confident individuals are more eager to receive performance feedback and women are less self-confident than men, they may demand less performance feedback.

Another reason men and women might differ in their demand for performance feedback could be anticipated responses to feedback. In studies that have considered exogenously provided performance feedback, some researchers have found that women update their beliefs

more conservatively than men do (Coutts 2019, Möbius et al. 2022), particularly in male-typed domains (Coffman et al. 2023a). A slew of recent work has documented that failure or negative feedback seems to deter women more so than men (Gill and Prowse 2014, Brown et al. 2019, Buser and Yuan 2019, Fang et al. 2021, Ellison and Swanson 2023, Pereda et al. 2023, Wasserman 2023, Kang et al. 2024). Similarly, Shastry et al. (2020) find that men are more likely to explain away negative feedback to luck; consistent with this finding, Coffman et al. (2023b) show that women hold more pessimistic beliefs about their abilities after receiving negative feedback compared with men. These findings raise the question of whether men and women, who differ in their responses to feedback, will also vary in their demand for feedback. If women anticipate more negative reactions to bad news, could this lead to lower demand for performance feedback?

Hypothesis 1. *Women demand performance feedback less than men.*

Hypothesis 1a. *Women are more pessimistic than men about their own performance.*

Hypothesis 1b. *Beliefs of own performance are positively related to demand for performance feedback.*

Our study designs allow us to test Hypothesis 1 as well as the subhypotheses that inform this prediction. Assuming evidence in favor of Hypothesis 1, 1a, and 1b, we expect that controlling for beliefs of own performance would help to explain any observed gender gap in the demand for performance feedback.

In addition, for both of our studies, we are interested in testing whether a gender gap in the demand for feedback is expected. To do so, we will describe our study designs to a separate set of subjects and elicit their forecasts of men and women's preferences for feedback. We hypothesize that these forecasters anticipate that women are more feedback averse than men.

Hypothesis 2. *Participants forecast that women demand performance feedback less than men.*

Finally, our last hypothesis is specific to study 2. Given recent evidence that women may anticipate being discriminated against in evaluations of their capabilities and take action to prevent such discrimination (e.g., Alston 2019, Exley et al. 2024), we are interested in studying whether women disproportionately shy away from performance feedback when feedback is provided by an individual who can discriminate on the basis of gender.

Hypothesis 3. *The possibility of gender discrimination in the performance evaluation leads women to decrease their demand for feedback more so than men.*

3. Demand for Feedback on a Cognitive Test

3.1. Experimental Design: Feedback Study

In the feedback study, participants take a cognitive test, report their beliefs of how they performed, and make decisions about whether they would like to receive accurate, objective feedback on how they performed.

3.1.1. Cognitive Test and Treatment Variation.

Participants have five minutes to answer 30 questions on arithmetic reasoning, assembling objects, math, general science, and mechanical comprehension, drawn from the Armed Services Vocational Aptitude Battery (ASVAB).[4] We intentionally chose questions from stereotypically male-typed domains. This increases the extent to which our results are likely to speak to real-world settings of interest where gender gaps are largest, such as in STEM fields. Each question is multiple choice. Participants receive \$0.10 per correct answer and \$0 for skipped or incorrect answers. As we detail later in this section, we add a stochastic component to the final payment so that participants cannot infer their performance from their earnings.

We randomize participants into either an easy or a hard version of the test in a between-subjects design. We exogenously manipulate the difficulty of the test in study 1 to examine whether beliefs of own performance causally impact the demand for feedback. This rationale builds on previous findings that beliefs about ability depend upon the difficulty of the task; specifically, individuals' confidence drops as the difficulty of the task increases (Moore and Healy 2008, Bordalo et al. 2019).[5] Assuming that randomly assigned difficulty level of the test does indeed impact beliefs of own performance, we can use this treatment assignment as an instrument for (over)confidence.

3.1.2. Prior Beliefs.

After completing the cognitive test, participants report their beliefs about absolute and relative performance and their degree of confidence in those beliefs. Participants first indicate how many questions they believe they answered correctly, receiving \$0.10 if their guess is correct and \$0 otherwise. Subjects then indicate how sure they feel about their guess on a 1–5 scale ranging from "not sure at all" to "completely sure." We choose this qualitative scale in hopes that it produces less measurement error among this population than a fully incentivized probabilistic elicitation.[6] Participants then indicate how they think they ranked relative to nine other randomly drawn study participants who completed the same test, receiving \$0.10 if their guess is correct and \$0 otherwise. Finally, participants indicate how sure they feel about their performance rank guess, again on a 1–5 scale.[7]

3.1.3. Preferences over Feedback. Following the belief reports, we elicit participants' preferences for receiving feedback on their performance. We do this in two parts. In the first part, we ask participants to indicate how interested they are in learning the number of questions they answered correctly and their rank relative to the other randomly chosen nine participants (as a single bundle of information) on a 1–5 scale ranging from "not at all interested" to "extremely interested."[8] This is an unincentivized report because the answer to this question does not determine whether the participant receives the feedback. Our goal is to collect a simple, intuitive measure, unlikely to generate confusion, before continuing to incentivized measures.

In the second part of the elicitation, we inform participants that they have an opportunity to learn at the end of the session how many questions they answered correctly and how they ranked relative to the other randomly chosen nine participants and that they will now be presented with three questions. Their answer to one randomly selected question determines whether they learn this information. For each of the three questions, participants must make a choice between two options: receiving or not receiving feedback. We vary the price associated with each option across the question. We use real-effort task prices, rather than monetary prices, to avoid potential "house money" effects. In particular, we ask participants to complete sliders (Gill and Prowse 2012, Araujo et al. 2016).[9] To familiarize participants with sliders, we required participants to complete two sliders before they advance to the three-question elicitation.

In question 1, the choice is between receiving or not receiving the information, with no real-effort price attached to either choice. In question 2, the choice is between receiving the information and completing 10 sliders or not receiving the information. In question 3, the choice is between receiving the information or not receiving the information and completing 10 sliders. Therefore, relative to question 1, question 2 adds a real-effort cost to acquire the feedback, and question 3 adds the same cost to avoid the feedback. The three questions are presented one at a time, on separate pages. Question 1 is always presented first, and the order of questions 2 and 3 is randomized across participant.

3.1.4. Exit Questionnaire. After reporting their preferences for receiving feedback, participants provide their year of birth, gender, race, region of residence, and whether they attended high school in the United States. Participants also indicate their beliefs about average gender differences in performance across all participants who completed the same test by choosing one of the following options in an unincentivized manner: on average, (i) women answered at least three more questions correctly than men, (ii) women answered one or two more questions correctly than men, (iii) women and men answered correctly the same number of questions, (iv) men answered one or two more questions correctly than women, and (v) men answered at least three more questions correctly than women.

In this final part of the experiment, we also collect participant perceptions of how informative they expect feedback to be. We ask participants to imagine they were informed that they performed *better* than they expected and to indicate, on a scale from 1 to 10, how much such feedback would (i) influence their own evaluation of their performance, (ii) give them information on their cognitive ability generally, and (iii) give them information on their capabilities in other aspects of life. Participants also answer the same three questions under the assumption that they were informed that they performed *worse* than they expected. We randomize which block of three questions (better than expected or worse than expected) participants see first. Whereas our design is intended to minimize the instrumental value of feedback, it is possible that subjects nevertheless anticipate that the feedback will be useful outside the study. We include these questions to capture participant perceptions of this form of instrumental value of the feedback.

3.1.5. Provision of Feedback. Following the exit questionnaire, participants who were selected to receive feedback based upon their choices learn their absolute and relative performance and complete any necessary sliders. We ask participants who receive feedback to type in the information they receive back to us on that same feedback screen; we inform them of this protocol at the time of their decisions. This ensures that participants make their choices over feedback knowing that they cannot avoid the feedback if they choose to receive it. Finally, all participants learn their total earnings in the study, and the session concludes.

3.1.6. Implementation. We conducted the study in June 2020 on the Amazon MTurk platform. A total of 995 subjects completed the study, with 502 assigned to the easy version and 493 to the hard version of the test. All participants received a fixed payment of \$2.50 plus a bonus payment that was divided into two components. The first component corresponded to their performance on the cognitive test and the accuracy of their beliefs, as detailed above. The second component was determined by a uniform random draw from [\$0, \$3] in increments of \$0.10. At the end of the session, we informed participants of their total earnings, but not of the breakdown of their earnings by components. We included the random earnings component to ensure that participants could not infer their absolute performance from their earnings, which would have diminished the value of receiving (or avoiding)

feedback. We explained this feature to participants at the beginning of the study and again during the elicitation of preferences over feedback. The study lasted 15–20 minutes and was open only to MTurk workers 18 years of age or older, with IP addresses located in the United States, at least 100 previous HITs completed on MTurk, and approval ratings of at least 95%. Participants had to pass several comprehension and attention checks distributed throughout the session to complete the study. We preregistered the study before data collection (Coffman and Klinowski 2020a).

3.2. Experimental Design: Forecast Study

After conducting the feedback study, we elicited beliefs about participants' demand for feedback from a separate sample of subjects. We call this the forecast study. We first elicit their demographic information (gender, age bracket, and region of residence) and then inform them about the feedback study. Subjects spend at least two minutes viewing the cognitive test, though they do not have to answer the test questions. Subjects are randomized into seeing either the easy or hard version of the test. We then describe to the subjects how we elicited the previous participants' demand for feedback on their performance on the test, and we familiarize subjects with the real-effort task prices by asking them to complete two sliders.

We elicit beliefs of the previous participants' demand for feedback. We elicit beliefs about male and female participants separately, asking, in each case, three questions. First, we ask subjects to guess how many out of 100 (male/female) participants chose to receive feedback when the price to receive feedback was zero and then to guess how many out of 100 (male/female) participants chose to receive feedback when it cost 10 sliders to receive feedback. Finally, we ask subjects to guess how many out of 100 (male/female) participants chose to receive feedback when it cost 10 sliders to avoid feedback. After subjects provide their guesses for these three questions for one gender, we ask the same three questions for the other gender. We randomize the order of the gender across subject and show the set of three questions is always in the same order within gender. Subjects receive a bonus of \$0.25 if one of their six guesses, randomly chosen, is within five percentage points (pp) of the correct answer.

Finally, we directly elicit beliefs about gender differences in the previous participants' willingness to receive feedback on their performance on the cognitive test and on tasks more generally. We do this by asking two unincentivized, Likert-scale questions: (i) "Overall, for the task participants completed in the previous study, how would you describe differences in men's and women's preferences for finding out how they performed?", and (ii) "Thinking more generally—not just for the task participants completed in the previous study—how would you describe gender differences in preferences for finding out how they performed in educational and professional settings, such as tasks in school and work?". For each question, subjects must choose one of the following options: (i) men want to find out how they performed much more than women, (ii) men want to find out how they performed somewhat more than women, (iii) there is no gender difference in interest in finding out how they performed, (iv) women want to find out how they performed somewhat more than men, and (v) women want to find out how they performed much more than men.

We conducted the study in November 2022 on the Prolific platform. A total of 982 subjects completed the study, with 492 of them assigned to see the easy version and 490 to the hard version of the test. All subjects received a fixed payment of \$4 plus a bonus for guessing correctly as described above. The study lasted 10–15 minutes and was open only to Prolific participants 18 years of age or older with IP addresses located in the United States, at least 100 completed studies on the platform, and an approval rate of at least 95%. We preregistered the study before data collection (Coffman and Klinowski 2022). Note that during this study, other subjects were randomized into treatments that elicited beliefs about demand for feedback on performance in the interview, as described in Section 4.

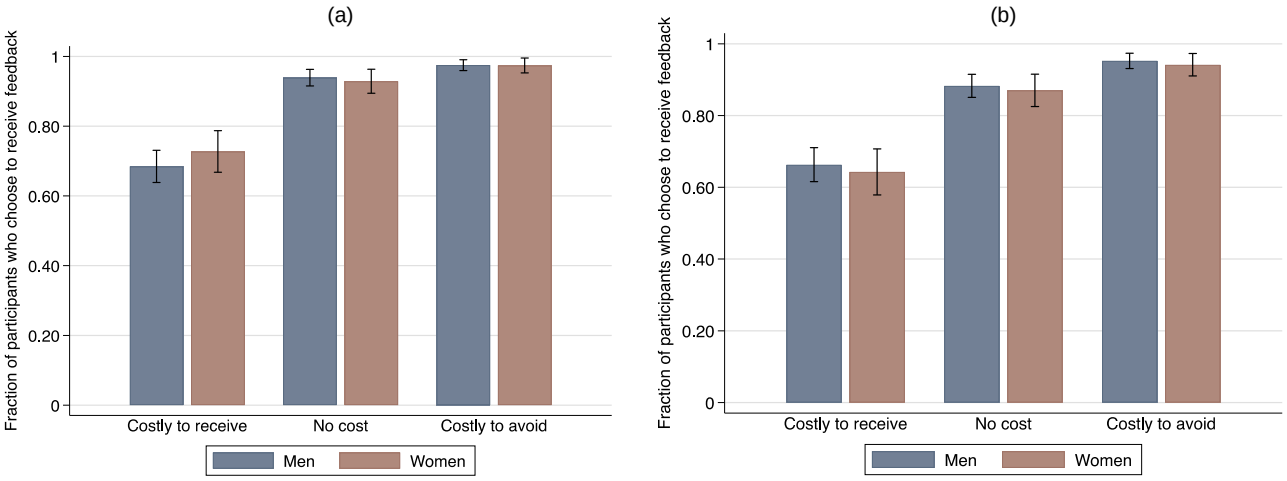
3.3. Results: Feedback Study

3.3.1. Descriptive Statistics. The sample consists of 350 women and 645 men in the feedback study and 463 women, 489 men, and 30 who identified as neither man nor woman in the forecast study.[10] Table A.1 in Online Appendix A provides descriptive statistics, and Table A.2 in Online Appendix A shows that demographic characteristics are largely balanced across treatments (easy and hard versions of the test) within the study. As expected, participants answer significantly more questions correctly on the easy version than the hard version of the test (9.4 versus 7.2, $p < 0.001$). There are no gender differences in test scores in either treatment (Table A.1 in Online Appendix A).[11]

3.3.2. Demand for Feedback. We start by examining the participants' preferences for feedback. We present results in this section for the 860 subjects with monotonic preferences, which constitute 86.4% of the 995 subjects who completed the cognitive test feedback study. Failure of monotonicity does not vary significantly across treatment ($p = 0.869$) or gender ($p = 0.287$) and is positively correlated with finding the instructions difficult and with lower test scores, suggesting that nonmonotonicity is most likely related to confusion (Table A.3 in Online Appendix A).

Overall, we find substantial demand for feedback. Looking first at the qualitative measure asking how

Figure 2. (Color online) Demand for Feedback on the Cognitive Test



Notes. (a) Easy version. (b) Hard version. Observations from the cognitive test feedback study. Sample restricted to subjects with monotonic preferences for feedback over all prices. Whiskers indicate 90% confidence intervals.

interested subjects were in learning about how they performed, the average is 3.78 on the 1–5 scale.[12] Women report greater interest in receiving feedback: 3.73 for men and 3.87 for women ($p = 0.093$). This gender difference is concentrated within the easy version. On the easy version, men report, on average, 3.70, whereas women report, on average, 4.03 ($p = 0.004$). On the hard version, men report on average 3.75, whereas women report, on average, 3.71 ($p = 0.702$). The difference in difference is significant in a regression ($p = 0.026$ without demographic controls, $p = 0.051$ with demographic controls).

Figure 2 shows the share of men and women who choose to receive feedback for each of the three price-list items and two treatment conditions. Demand is high, and there is no gender difference in this demand. More than two-thirds of our participants choose feedback even when it is costly. For both the easy and hard tests, demand for feedback declines as feedback becomes more costly to receive ($p < 0.001$ for all pairwise comparisons), suggesting attentiveness and understanding among participants. More critically, there is no significant gender gap in demand for feedback at any price in either treatment.

To examine in more detail the gender gap in the demand for feedback, we use ordinary least squares (OLS) regressions to estimate the probability that the participant chooses to receive costly feedback on a female indicator and additional covariates as indicated in Table 1. We focus on the choice of receiving feedback when it is costly to do so because the variability is greatest for this outcome (as we note below, results are similar for the choices of receiving feedback when it is costly to avoid it and when there are no costs for either option). Table 1, columns (1) and (2) show that demand is not significantly different across treatment or gender, irrespective of the inclusion of demographic controls (there is no significant interaction effect between treatment and gender; Table A.4 in Online Appendix A, column (1)).[13] The result holds regardless of the order of the elicitation of preferences over feedback. The 95% confidence interval of the female coefficient in column (2) is (−0.038, 0.092). Column (3) shows that these results are unchanged when we control for test performance. On average, performing better on the test (lower rank) is associated with being less likely to choose to receive feedback; this is seen by the positive coefficient on actual rank, which is a continuous variable from 1 to 10, where 1 is the best decile and 10 the worst decile of performance within the treatment condition. In sum, across specifications 1–3 in Table 1, we find that women are no less likely than men to demand costly feedback. This result holds when there is no cost to receive or avoid feedback (Table A.5 in Online Appendix A) and when avoiding feedback is costly (Table A.6 in Online Appendix A). These robustness checks help to alleviate concerns that gender differences in the perceived or real costs of completing sliders drive our results.[14]

Result 1. *Contrary to Hypothesis 1, we cannot reject that men and women have the same demand for performance feedback.*

Next, we explore our subhypotheses, investigating the relationship between beliefs of performance and demand for costly feedback across gender.

3.3.3. The Role of Beliefs. On average, participants overestimate their scores. Average beliefs of absolute score are 11.4 on the easy version and 9.6 on the hard

Table 1. Demand for Costly Feedback on the Cognitive Test

	(1)	(2)	(3)	(4)	(5)
Female	0.011	0.027	0.031	0.041	0.053
	(0.033)	(0.033)	(0.033)	(0.033)	(0.034)
Hard	−0.044	−0.048	−0.048	−0.033	−0.028
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Actual rank			0.014***	0.016***	0.015***
(1: best, 10: worst)			(0.006)	(0.006)	(0.006)
Belief of test score				0.005**	
				(0.002)	
Certainty in test score				0.023	
				(0.014)	
Belief of rank					−0.023****
(1: best, 10: worst)					(0.006)
Certainty in rank					0.029**
					(0.014)
Controls	N	Y	Y	Y	Y
N	860	860	860	860	860
R^2	0.0024	0.0262	0.0338	0.0448	0.0541

Notes. Observations from the cognitive test feedback study. Coefficient estimates from OLS regressions of the probability that the participant chooses to receive feedback when it is costly to receive it. Controls are age, race, region of residence, high school in the United States, and the order of the three questions that elicit preferences for feedback. Sample restricted to subjects with monotonic preferences for feedback. Robust standard errors are in parentheses.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; ****$p < 0.001$.

version ($p < 0.001$). Figure A.1, (a) and (b), in Online Appendix A plots beliefs about test score as a function of test score. Conditional on test performance, women's beliefs are significantly less optimistic than men's on the both the easy and hard versions of the test ($p = 0.056$ and $p = 0.023$, respectively; Table A.7, columns (1) and (2) in Online Appendix A).

In terms of beliefs of relative performance, participants, on average, rank themselves about in the middle of the pack, with an average rank of 5.8 on the hard test and 5.2 on the easy test, where a rank of 1 is best and 10 is worst. On average, women are significantly less optimistic than men about their relative performance (Figure A.1, (c) and (d) in Online Appendix A). Conditional on true rank, women's beliefs of rank are approximately 0.8 ranks lower than men's ($p < 0.001$; Table A.8, column (3) in Online Appendix A). Both the easy and the hard tests produce this gender gap ($p < 0.001$ and $p = 0.005$, respectively; Table A.8, columns (1) and (2) in Online Appendix A), with no significant difference in the gap across treatments (n.s. interaction term in column (4) of Table A.8 in Online Appendix A).[15] Overall, our evidence is consistent with Hypothesis 1a.

Result 1a. *Women hold significantly more pessimistic beliefs about their performance than men.*

Also consistent with past evidence, beliefs of both absolute and relative performance are positively predictive of demand for feedback. Returning to the estimation of the decision to receive costly feedback in Table 1, column (4) includes as covariates beliefs of absolute performance and certainty of these beliefs, and column (5) includes, instead, beliefs of relative performance and certainty of these beliefs. In both specifications, conditional on actual rank, more optimistic participants have significantly greater demand for feedback ($p = 0.018$ for beliefs of absolute performance and $p < 0.001$ for beliefs of relative performance). This is true for both men and women (Table A.4, columns (3) and (4) in Online Appendix A). Certainty of beliefs is also predictive of demand for feedback. Greater certainty is associated with greater demand, directionally so for beliefs of absolute performance ($p = 0.107$; Table 1, column (4)) and statistically significantly so for beliefs of relative performance ($p = 0.046$; Table 1, column (5)).[16] A regression that drops actual rank from Table 1, column (5) does not change our conclusions for the remaining coefficients.

Finally, we examine whether self-confidence causally affects the demand for feedback by using random assignment to treatment to instrument for beliefs of relative performance. Treatment successfully manipulated confidence: beliefs of relative performance are significantly more pessimistic in the hard version conditional on true rank ($p < 0.001$; Table A.8, column (3) in Online Appendix A). Directionally, we find that more optimistic participants are more likely to demand feedback, as the two-stage least squares (2SLS) estimate of the effect of beliefs on the demand for feedback is −0.064 ($p = 0.165$; Table A.12, column (2) in Online Appendix A), which is more than twice as large as the OLS estimate of −0.024 ($p < 0.001$; Table A.12, column

(1) in Online Appendix A) but much less precisely estimated.[17] Taken as a whole, our evidence supports Hypothesis 1b.

Result 1b. *Beliefs of own performance are positively related to demand for performance feedback.*

In sum, women are significantly less confident of their own performance than men, and on average, more self-confident participants are more likely to demand feedback. Despite this, we do not observe a gender gap in demand for feedback.[18]

3.4. Results: Forecast Study

In this section, we examine the forecasts made by a separate set of subjects about the demand for feedback on the cognitive test. We restrict the forecaster sample to the 744 participants who have monotonic forecasts over nonzero prices (participants who forecasted that more subjects opt to receive feedback when it is free than when it is costly), which constitute 75.8% of the 982 participants in the cognitive test forecast study. Results are similar for the full sample of participants and for forecasts of the choice of receiving feedback at zero price (Tables A.13 and A.14 in Online Appendix A).[19]

Table 2 estimates the forecasted probability that a participant opts for feedback when they must complete two sliders to receive it. Recall that we ask participants to make predictions about how many out of 100 men (women) would demand feedback at this price. We translate this into probability, with coefficients in hundreds of percentage points, to allow for easier comparability with our results from the feedback study. Column (1) shows results from a within-subject analysis, regressing the forecasted probability of demanding feedback on an indicator that the forecast is about women and subject fixed effects. We estimate that participants forecast 58.5% of men and 55.0% of women opt for feedback, a difference of 3.5 pp ($p < 0.001$). Note that this forecasted difference is outside of our 95% confidence interval around the estimated gender difference; comparing the analogous specification 2 from Tables 1 and 2, we estimate a forecasted gender difference of −0.044, outside of our confidence interval for the observed gender difference, (−0.038, 0.092).

Table 2. Forecasts of the Demand for Costly Feedback on the Cognitive Test

	(1)	(2)	(3)
Female	−0.035****	−0.044**	−0.044**
	(0.007)	(0.017)	(0.017)
Hard		−0.027	−0.027
		(0.017)	(0.017)
Female forecaster			0.030*
			(0.018)
Within subjects	Y	N	N
Across subjects	N	Y	Y
Controls	N	Y	Y
Mean	0.568	0.564	0.564
N	1,488	744	744
R^2	0.8286	0.0218	0.0257

Notes. Observations from the cognitive test forecast study. Coefficient estimates from OLS regressions of the forecasted probability of opting for feedback when feedback is costly. Controls are age bracket and region of residence. Sample restricted to the 744 subjects with monotonic forecasts over nonzero prices for feedback. Robust standard errors are in parentheses.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; ****$p < 0.001$.

Column (2) shows results from an across-subjects analysis, restricting the sample to only the first forecasts provided by the participant (recall that subjects are randomized into being asked first about men or first about women) and controlling for demographics (the forecast provider's age bracket and region of residence). We continue to estimate a gender gap in this specification, with participants forecasting that women are 4.4 pp less likely to demand feedback ($p = 0.011$). Column (3) shows that controlling for the forecaster's gender does not eliminate the forecasted gender gap. The forecasted gender gap in demand for feedback is indistinguishable across the hard and easy versions of the cognitive test, and both men and women forecast that women have lower demand for feedback (Table A.15 in Online Appendix A).

Recall that participants also answered two qualitative Likert questions that elicited their beliefs of gender differences in demand for performance feedback on the cognitive test and on tasks more generally. For both questions, the median answer is that men seek feedback somewhat more than women. If we code the answers on a scale {−2, −1, 0, 1, 2}, with positive (negative) values indicating that men (women) seek relatively more feedback and zero indicating no gender difference, a Wilcoxon signed-rank test rejects equality of the distribution around zero ($p < 0.001$) for both questions. The mean answer is 0.39 ($p < 0.001$) for the cognitive test question and 0.28 ($p < 0.001$) for the general question.[20]

Result 2. *Participants forecast that women demand performance feedback less than men, both in terms of the cognitive test study and more generally.*

Thus, women in our feedback study are no less eager for performance feedback than men, yet both men and women in the forecast study anticipate that women are more feedback averse.

In these studies, feedback is objective and private. A computer scores the test, and feedback simply involves privately viewing that score and relative rank. However, in many settings of interest, feedback is likely to be substantially more subjective and often provided by another individual. Is it possible that women become less eager to receive feedback in these more subjective

settings, particularly when there is a possibility of gender discrimination in the feedback itself? We explore this in a second set of studies.

4. Demand for Feedback in an Interview Setting

In conducting a second study, we provide additional tests of our main hypotheses (Hypotheses 1 and 2): do women demand performance feedback less than men, and do individuals believe that women demand performance feedback less than men? In addition, study 2 varies across participants whether it is possible for the feedback provider—an HR professional in this case—to discriminate on the basis of gender.

4.1. Experimental Design: Feedback Study

The feedback study consists of two sessions that occur three weeks apart. In session 1, participants answer common job interview questions about their life achievements and personality. We then hire HR professionals to rate the answers given by a random subset of participants. We manipulate whether the HR professionals observe the gender of the participants when evaluating their answers, which allows us to test whether the possibility of gender-based discrimination by the HR professional impacts preferences for feedback. After obtaining the ratings from the HR professionals, we invite participants back and conduct session 2. In session 2, participants are informed of the HR rating step that occurred between sessions and make decisions about whether they would like to receive feedback on how they ranked relative to other participants based on the ratings assigned by the HR professionals.

4.1.1. Session 1. Session 1 starts by asking participants for three pieces of demographic information: sex, age bracket (18 to 30, 31 to 50, or 51+), and region of residence (Northeast, South, Midwest, or West). We collect this information of the participant up front so that we can reveal it later to the HR professional. Whereas we are interested in the sex of the participant, we also collect the age bracket and the region of residence in order not to focus attention on sex and thus minimize any potential priming or experimenter demand effects (De Quidt et al. 2019). After providing these demographics, participants proceed to the task. They answer three questions that are commonly asked in job interviews to assess candidates. They have five minutes to answer each question by typing their answers on their computer or devices. Participants are informed that they may be disqualified from participating in the second session of the study if they answer any of the interview questions in less than 60 words, submit an answer in two minutes or less, or navigate away from the screen during the interview. We also let participants know that copy-pasting has been disabled. This helps to ensure that participants give honest answers and to establish common knowledge of this fact, making the feedback about relative performance more meaningful to participants. The three interview questions are, "What is something you have achieved that you are proud of, and why?", "Describe a difficult task you were faced with and how you addressed it," and "What are you passionate about, and why?". The questions appear one at a time on separate screens. Following the three interview questions, participants are reminded that they will be invited to a second session in three weeks.

4.1.2. Evaluation by HR Professionals. In the three-week interim between sessions 1 and 2, we hire two HR professionals from the platform Upwork to rate the answers to the interview questions of a subset of participants in session 1. We randomly assign participants in session 1 to either a blind or a nonblind condition in a between-subjects design, with a different HR professional assigned to each condition. In each condition, we randomly choose 10 participants to have their answers to the interview questions evaluated by the HR professional. In the blind condition, the HR professional is shown the answers to the interview questions for each of the 10 participants, but not their demographic information. In the nonblind condition, the HR professional is shown both the answers to the interview questions and the demographic information (sex, age bracket, and region of residence) of the 10 participants evaluated. Other than this difference, the evaluation sessions are identical across conditions.

During the evaluation, we instruct the HR professional to rate each of the 10 participants on four different traits, based on the participant's answers to the three interview questions. The four traits are intellectual curiosity, a tendency to strive for achievement, assertiveness, and tolerance to stress. The HR professional assigns each participant a score from 1 to 10 on each trait that indicates the extent to which the participant's answers demonstrate the trait.

4.1.3. Session 2. Three weeks after completing session 1, participants return for session 2, in which we inform them that there was some chance that their answers to the interview questions have been evaluated by an HR professional, and we give them an opportunity to receive feedback on how they ranked on the evaluation relative to other participants.

4.1.3.1. Preliminary Instructions. At the beginning of session 2, we remind participants that in session 1, they provided their demographic information (sex, age bracket, and region of residence) and answered three questions commonly asked in job interviews.

We inform participants that an HR professional with experience evaluating job candidates rated the answers given by 10 randomly selected participants, assigning each of them a score from 1 to 10 on four traits that are generally valued by employers: intellectual curiosity, a tendency to strive for achievement, assertiveness, and tolerance to stress (we include a brief definition of each trait). We tell participants that for each participant rated, we have constructed an "interview score" by averaging the subject's score across the four traits.

In the blind condition, we (truthfully) inform participants that the HR professional saw the answers to the interview questions and no other information about the participants. In the nonblind condition, we (truthfully) inform participants that the HR professional saw both the answers to the interview questions and the sex, age bracket, and region of residence of each participant. This treatment variation allows us to study whether anticipation of potential gender-based discrimination by the HR professional leads to gender differences in the demand for feedback. To sharpen the treatment, we mention several times throughout this preliminary information stage what information was seen by the HR professional. We also include understanding questions, one of which requires participants to indicate correctly what information was seen by the HR professional before they can advance to the next stage of the session.

4.1.3.2. Prior Beliefs.

After participants receive the preliminary information, we elicit their beliefs about their rank on the interview score relative to the other nine participants randomly selected to have their interview answers rated by the HR professional. Participants report their believed rank and, if indeed they were randomly selected to be ranked by the HR professional, they receive \$10 if their guess is correct and \$0 otherwise. We then elicit the precision of this belief by asking participants to indicate how sure they feel about their guess on a 1–5 scale.

4.1.3.3. Preferences over Feedback.

We elicit participants' preferences for receiving feedback on their relative performance in the interview in two parts.[21] In the first part, we ask participants in an unincentivized fashion to indicate how interested they are in learning their rank on the interview score on a 1–5 scale ranging from "not at all interested" to "extremely interested." In the second part, we measure participants' willingness to pay a real-effort cost (completing sliders) to receive and avoid feedback, as in study 1. However, this time, we include a more granular and wider-ranging real-effort price list.

After familiarizing participants with the slider task, we present participants with the price list as a sequence of 11 questions that appear one at a time on separate pages. We randomly select one of these 11 questions to determine their outcome, conditional on having indeed been ranked by the HR professional. The first question is always a choice between (i) being told the rank on the interview score, or (ii) not being told the rank on the interview score. This question is followed by a block of five questions that involve a choice between (i) being told the rank on the interview score and completing X sliders, or (ii) not being told the rank on the interview score. This block is followed in turn by a second block of five questions that involve a choice between (i) being told the rank on the interview score, or (ii) not being told the rank on the interview score and completing X sliders. Within a block of five questions, X always increases from 2, 5, 10, 50, to 100. (We estimate that it would take participants seven to eight minutes on average to complete 100 sliders, which corresponds to about 35%–40% of the average session 2 completion time.) The order of the two five-question blocks is randomized.

4.1.3.4. Exit Questionnaire.

Following the preference elicitation, we instruct participants to imagine they received information that they ranked *better* than they expected and ask them to indicate on a 1–10 scale how much such feedback would (i) influence their own evaluation of their abilities in terms of the traits of intellectual curiosity, a tendency to strive for achievement, assertiveness, and tolerance to stress; (ii) lead them to change their beliefs about their ability to perform well on a job interview; and (iii) lead them to change their beliefs about their capabilities in other aspects of life. Participants also answer the same three questions under the assumption that they received feedback that they performed *worse* than they expected, and we randomize which block of three questions (better than expected or worse than expected) participants see first.

We then ask participants for race, educational attainment, current employment status, and whether they attended high school in the United States. Following the demographics questions, participants indicate their agreement on a 1–7 scale with the statement, "In the past, I have worried whether I have been treated or evaluated unfairly because of my sex." They also indicate whether they think in the future, when trying to find or keep a job, employers will treat or evaluate them (substantially less, slightly less, equally, slightly more, substantially more) favorably than others because of their sex. These two questions are designed to elicit participants' beliefs of past and future sex-based discrimination. Finally, participants indicate their beliefs about average gender differences in the interview score across all participants in the treatment condition by choosing one of the following options in an unincentivized manner: on average, (i) women obtained a much better interview score than men, (ii) women obtained a slightly

better interview score than men, (iii) women and men obtain equal interview scores, (iv) men obtained a slightly better interview score than women, and (v) men obtained a much better interview score than women.

4.1.3.5. Provision of Feedback.

At the end of the session, participants learn whether they had been randomly selected to have their answers to the interview questions evaluated, and if so, they receive information on their rank on the interview score depending on their answer to the elicitation question that was implemented. They also complete the corresponding number of sliders, if applicable. As in the cognitive test study, we require participants who receive feedback to type in the feedback they receive back to us on the same feedback screen, and we inform participants of this feature before they respond to the elicitation mechanism.

4.1.4. Implementation.

For sessions 1 and 2, we advertised the study on Amazon MTurk as an academic study involving two sessions, three weeks apart from each other, with a guaranteed payment of $1 for completing session 1 and an additional $5 for completing session 2. The larger fee for session 2 was intended to discourage attrition between sessions. The study was open to MTurk workers 18 years of age or older with IP addresses in the United States, at least 500 previous HITs completed on MTurk, and approval ratings of at least 95%.

We conducted session 1 in November 2020. Session 1 lasted approximately 20 minutes and included several understanding and attention checks. Our preregistered goal was to collect 1,500 observations that would be eligible for session 2. To meet this goal, we invited a total of 2,451 participants to session 1, yielding 1,515 participants who gave answers to the three interview questions that complied with the rules we established (at least 60 words per question, submitted in no less than two minutes, and typed without navigating away from the study page). Of these, we randomly assigned 759 to the blind condition and 756 to the nonblind condition. In each condition, we randomly selected 10 participants to have their answers evaluated by an HR professional. Note that session 1 is identical across treatments; participants do not receive treatment-specific instructions until session 2.

We collected the HR professional evaluations following session 1. We used Upwork to recruit HR professionals. We advertised the rating task as a one-time, one-hour job involving evaluating answers to a mock job interview given by 10 participants of an academic study, for a fee of $50. We opened the job to HR professionals with at least one year of experience in evaluating candidates in job interviews, and we selected two. They completed the evaluation session by providing their ratings via a Qualtrics survey.

We opened session 2 only to the 1,515 subjects who participated in session 1 and gave valid answers to the interview questions. We sent these individuals an invitation to participate and reminders of session 2 a few days prior to session 2. A total of 1,350 subjects participated in session 2. We included several attention and understanding checks throughout the session. Session 2 lasted approximately 20 minutes. We preregistered the interview preferences for the feedback study (see Coffman and Klinowski 2020b).

4.2. Experimental Design: Forecast Study

After conducting the feedback study, we elicited forecasts about the participants' demand for feedback from a separate sample of subjects. We call this the forecast study. From these new subjects, we first elicit their demographic information (gender, age bracket, and region of residence) and then inform them about the feedback study in its entirety and familiarize subjects with the real-effort task prices by asking them to complete two sliders before they can proceed.

We then elicit the subjects' forecasts of the previous participants' demand for feedback. We elicit forecasts about male and female participants separately, asking, in each case, three blocks of questions. The first block of questions consists of only one question, in which we ask subjects to guess how many out of 100 (male/female) participants chose to receive feedback when the price to receive feedback was zero. Then, the second block of questions consists of five questions, in which we ask subjects to guess how many out of 100 (male/female) participants chose to receive feedback when it cost 2, 5, 10, 50, and 100 sliders to *receive* feedback. Finally, the third block of questions consists of five questions, in which we ask subjects to guess how many out of 100 (male/female) participants chose to receive feedback when it cost 2, 5, 10, 50, and 100 sliders to *avoid* feedback. After subjects provide their guesses in these three blocks of questions for one gender, we ask the same three blocks of questions for the other gender. These blocks always appear in the same order (no price, positive price for receiving feedback, positive price for avoiding feedback). We randomize the order of which gender the participant provides guesses for first across subjects. Subjects receive a bonus of $0.25 if one randomly selected guess (of the 22 total guesses, 11 for each gender) is within five pp of the correct answer.

Finally, we elicit forecasts about gender differences in demand for feedback using two unincentivized qualitative questions, one asking for beliefs about the gender gap in demand for performance feedback on the interview task specifically and the other asking for beliefs about the gender gap in demand for performance feedback more broadly.

This study was run as a branch of the same experiment used to elicit forecasts about demand for performance

feedback on the cognitive test (see Section 3), conducted in November 2022 on the Prolific platform. A total of 970 subjects completed the interview version of the study, with 466 of them assigned to the HR-blind treatment and 504 to the HR-nonblind treatment.

4.3. Results: Feedback Study

4.3.1. Descriptive Statistics. Table B.1 in Online Appendix B shows descriptive statistics of the sample for the feedback study, which consists of the 1,350 subjects who completed session 2 and thus provided their full set of demographic information and demand for feedback. These subjects constitute 89.1% of the 1,515 subjects who completed session 1 and were invited to participate in session 2. Observable demographics are balanced across treatment conditions (Table B.2 in Online Appendix B). Attrition from session 1 to session 2 was directionally smaller in the blind condition than in the nonblind condition (10% versus 12%, chi-squared test $p = 0.153$) and directionally smaller for women than for men (10% versus 12%, chi-squared test $p = 0.203$). Since all participants underwent identical procedures up to the start of session 2, any differential attrition across treatment is likely due to chance. Moreover, because subjects were not informed of the opportunity to receive feedback in session 1, attrition is unlikely to be directly related to preferences for feedback.

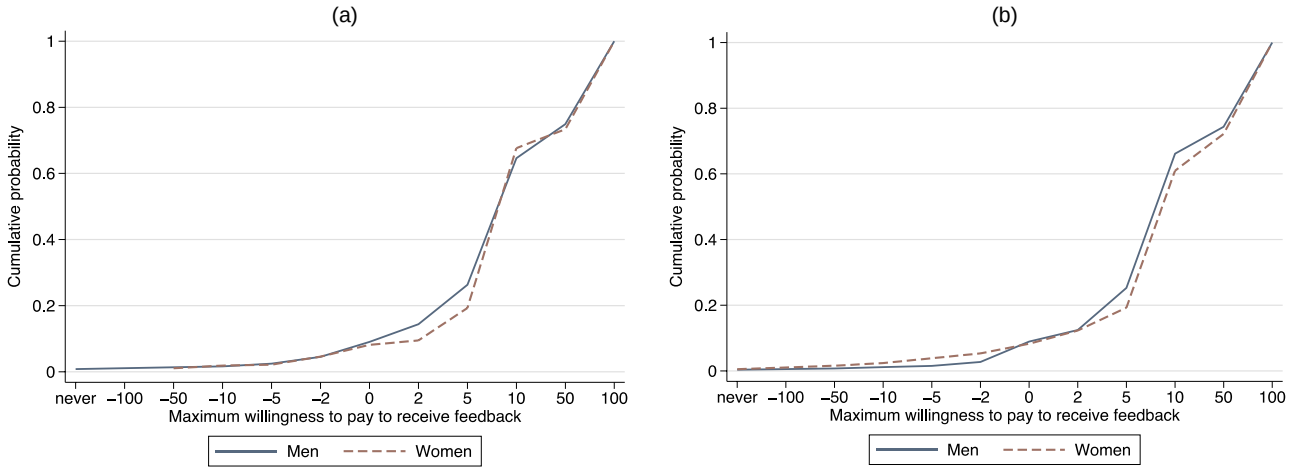
Unlike the cognitive test, the interview task does not produce an obvious, objective measure of performance. We constructed an objective measure of performance using the IBM Watson Personality Insights AI, a commercial artificial intelligence (AI) program that generates a personality profile from text. Using the participant's answers to the three interview questions, the AI outputs a personality profile consisting of a score from 0 to 1 on each Big Five personality trait and each facet component of each trait, where a higher score indicates that the participant exhibits the trait or facet to a larger extent. We extract the participant's scores on the facets of intellectual curiosity, assertiveness, striving for achievement, and tolerance to stress. We take the average of these four scores as the participant's objective score on the interview. We use this objective score as a control for performance in many of the analyses below, but note that participants are not made aware of this measure of performance, nor is it offered as feedback.[22]

4.3.2. Demand for Feedback. We present results in this section for 1,242 subjects with monotonic preferences for feedback over all prices, which constitute 92.0% of the 1,350 participants who completed session 2 of the feedback study. Failure of monotonicity is significantly correlated with being male, finding the instructions difficult, having lower AI score, and reporting lower interest in receiving feedback in the unincentivized elicitation (Table B.3 in Online Appendix B).

As in the cognitive test study, we see substantial demand for feedback. The average response to the qualitative measure of interest in receiving feedback on the 1–5 scale is 4.12. There is no significant treatment difference in this response: 4.08 in the nonblind condition and 4.16 in the blind condition ($p = 0.184$). As in the cognitive test study, women report more interest in receiving feedback: 4.05 for men and 4.17 for women ($p = 0.058$). There is no significant treatment-gender interaction ($p = 0.615$ from an OLS regression with no controls).

Figure 3 shows the cumulative distribution of the maximum a subject is willing to pay to receive feedback.[23]

Figure 3. (Color online) Maximum Willingness to Pay for Feedback on the Interview



Notes. (a) Nonblind condition. (b) Blind condition. Observations from the interview feedback study. Sample restricted to 1,242 subjects with monotonic preferences for feedback over all prices, which constitute 92.0% of the subjects who completed session 2.

Table 3. Maximum Willingness to Pay for Feedback on the Interview

	(1)	(2)	(3)	(4)	(5)
Female	3.523	0.713	−0.577	0.377	1.312
	(3.173)	(3.179)	(4.478)	(3.159)	(3.146)
Blind	2.527	2.859	1.359	2.973	2.492
	(3.118)	(3.074)	(4.837)	(3.065)	(3.043)
Female × Blind			2.544		
			(6.314)		
Average AI z-score				4.296*	4.571*
				(2.365)	(2.349)
Word count				0.037**	0.043***
				(0.016)	(0.016)
Belief of rank (1: best, 10: worst)					−1.481
					(0.931)
Certainty in rank					5.791****
					(1.663)
Controls	N	Y	Y	Y	Y
N	1,309	1,309	1,309	1,309	1,309

Notes. Observations from the interview feedback study. Coefficient estimates from interval regressions of the maximum willingness to pay to receive feedback when receiving feedback is costly (i.e., when the price to receive feedback is 2, 5, 10, 50, or 100 sliders). Controls are age, race, region of residence, educational attainment, high school in the United States, currently looking for a job, and the order of the block of five questions that elicit willingness to pay to receive or to avoid feedback. Sample restricted to 1,309 subjects with monotonic preferences for feedback over positive prices, which constitute 97.0% of the subjects who completed session 2. Robust standard errors are in parentheses.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; ****$p < 0.001$.

The majority of participants (91.5%) are willing to pay a strictly positive amount. The median and modal willingness to pay is 10 sliders. Strikingly, 26.5% of participants chose to receive feedback irrespective of the price. More central to our investigation, Figure 3 shows no large differences in the distributions across genders.

As in the cognitive test, we test for a gender gap in the demand for feedback using regressions and focusing on the choice of receiving feedback when it is costly to do so—that is, prices of 2, 5, 10, 50, and 100 sliders to receive feedback—because variability is greatest over positive prices. We use interval regressions to estimate the participant's maximum willingness to pay for feedback over positive prices on a female indicator and additional covariates as indicated in Table 3, restricting the sample to 1,309 subjects with monotonic preferences for feedback over positive prices, representing 97.0% of the 1,350 subjects who completed session 2. Consistent with the cognitive test study, we find no significant evidence of gender differences in demand for feedback and, if anything, directionally greater demand by women in most specifications. There are no gender differences in the maximum willingness to pay regardless of inclusion of demographics controls (Table 3, columns (1) and (2)) and regardless of the order of the elicitation of preferences over feedback.[24] We estimate that women are willing to pay 0.71 sliders more than men for feedback, with a 95% confidence interval of (−5.52, 6.94). Put differently, we can reject that men are willing to pay 5.5 sliders or more for feedback than women at the $p = 0.05$ threshold.

The total number of words written across the three interview questions is predictive of greater willingness to pay ($p = 0.017$; Table 3, column (4)), perhaps reflecting an association between effort in answering the questions and greater demand for feedback. Participants with higher objective scores as determined by the AI have greater willingness to pay ($p = 0.069$; Table 3, column (4)).[25] Conditioning on these measures continues to produce a null gender gap.

Result 1 (replication). *Contrary to Hypothesis 1, we cannot reject that men and women have the same demand for performance feedback.*

We do not find that the possibility of sex-based discrimination in the performance evaluation impacts participants' demand for feedback. The coefficient on *Blind* in columns (1) and (2) in Table 3 reveals that, on average, there was no significant difference in demand for feedback across the two treatments. The insignificant interaction effect in column (3) reveals that women were no more responsive to the treatment variation than men.[26]

Result 3. *The possibility of sex-based discrimination in the performance evaluation does not significantly change demand for performance feedback among women or men.*

The observed relationship between beliefs and demand is directionally consistent with the findings in the cognitive test study. Women's beliefs are, on average, more pessimistic than men's, conditional on AI score rank (Table B.5 in Online Appendix B). The average

degree of certainty is 2.54 on the 1–5 scale, with no significant gender or treatment difference (Table B.6 in Online Appendix B). Participants who are more optimistic about their relative performance are willing to pay more for feedback, although the relation is not significant ($p = 0.112$; Table 3, column (5)). As in the cognitive test study, greater certainty of beliefs is significantly associated with greater demand ($p < 0.001$; Table 3, column (5)).[27] Conditioning on beliefs, we continue to estimate a null gender difference in the demand for feedback (Table 3, column (5)).

Result 1a (replication). *Women hold significantly more pessimistic beliefs about their performance than men.*

Result 1b (replication). *Beliefs of own performance are (directionally) positively related to demand for performance feedback.*

In sum, we do not find that women demand feedback less than men.[28] There are also no gender differences in the demand for feedback at zero price (Table B.4 in Online Appendix B). Furthermore, the possibility of sex-based discrimination does not significantly change demand for feedback among women or men.

4.4. Results: Forecast Study

We now turn attention to the forecasts made by a separate set of subjects about the demand for feedback on the interview. To mirror the analysis in the previous section, we use the forecasts about how many of 100 men (women) demand feedback at different prices to construct implied forecasts of maximum willingness to pay. We do this using only observations from the 841 participants who provided monotonic forecasts over positive prices, which constitute 86.7% of the sample who completed the interview forecast study. Table 4 presents the results from interval regressions that predict the implied forecasted maximum willingness to pay. Results are similar for the choice of receiving feedback at zero price in the full sample of participants (Table B.8 in Online Appendix B).

Table 4, column (1) performs a within-subject analysis, regressing the maximum willingness to pay on an indicator that the forecast is about women and subject fixed effects. We estimate that subjects forecast women are willing to pay 0.73 fewer sliders than men ($p < 0.01$). Column (2) performs an across-subjects analysis, using only the first set of forecasts the participant provided (recall that subjects are randomized into being asked first about men or first about women). In this specification, we estimate a larger forecasted gender gap: subjects forecast women are willing to pay 4.16 fewer sliders than men ($p = 0.018$). Thus, just as in our cognitive test study, respondents forecast that women's demand for feedback will be less than men's, whereas our observed gap points in the other direction. But in this case, the forecasted female-male difference of −4.16 does fall within our 95% confidence interval around the estimated female-male gap, (−5.52, 6.94). Column (3) shows that controlling for the forecaster's gender does not eliminate the forecasted gender gap. The forecasted gender gap in demand for feedback is indistinguishable across the blind and nonblind conditions, and both men and women forecast that women have lower demand for feedback (Table B.9 in Online Appendix B).

Table 4. Forecasts of the Maximum Willingness to Pay for Feedback on the Interview

	(1)	(2)	(3)
Female	−0.728***	−4.158**	−4.151**
	(0.224)	(1.764)	(1.777)
Blind		2.007	1.949
		(1.998)	(1.993)
Female forecaster			0.986
			(1.945)
Within subjects	Y	N	N
Across subjects	N	Y	Y
Controls	N	Y	Y
Mean	28.677	25.353	25.353
N	168,200	84,100	84,100

Notes. Observations from the interview forecast study. Coefficient estimates from interval regressions of the maximum willingness to pay to receive feedback when receiving feedback is costly (i.e., when the price to receive feedback is 2, 5, 10, 50, or 100 sliders). Controls are age bracket and region of residence. Sample restricted to the 841 subjects with monotonic forecasts about preferences for feedback over positive prices. Each subject generates a set of 100 observations corresponding to the subject's forecasts of the feedback choice of 100 males and a set of 100 observations corresponding to the subject's forecasts of the feedback choice of 100 females. Column (1) uses the full set of 200 observations per subject. Columns (2) and (3) use 100 observations per subject, corresponding to the forecast about the gender first elicited from the subject. Robust standard errors in parentheses in column (1) and clustered at the subject level in columns (2) and (3).

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; ****$p < 0.001$.

Finally, for our unincentivized, qualitative questions that elicited the forecasters' beliefs of gender differences in demand for performance feedback on the interview task and on tasks more generally, we find that forecasters believe women are less likely to demand performance feedback than men on the interview task specifically, as the median answer is that men seek feedback more than men, a Wilcoxon signed-rank test rejects equality of the distribution around zero ($p = 0.057$), and the mean answer is 0.098 ($p = 0.034$). However, for tasks more generally, forecasters anticipate no gender differences in demand for feedback; a Wilcoxon signed-rank test fails to reject equality of the distribution around zero ($p = 0.221$), and the mean answer is 0.034 (t-test $p = 0.435$). This is in contrast to what we found for the identical generalized question from forecasters assigned to the cognitive skills test version of the study. It is possible that the "broader tasks" that come to mind for participants

differ depending upon whether they have been thinking about the cognitive test versus the interview questions, but we cannot know for sure what drives this difference.

5. Discussion

Across two studies, we elicit individuals' demand for feedback on their performance on a task. We vary whether the task is a cognitive test or answering interview questions, whether the feedback is objective or subjective, and whether it is possible for the provider of the feedback to discriminate based on the participant's gender. Across these variations, we find that women demand performance feedback no less than men, and when we elicit forecasts from other individuals, we observe that they anticipate women being less eager to receive performance feedback than men in our setting. These findings are summarized in Table 5, which presents an analysis pooling observations from our two studies. Column (1) shows the estimated probability that a participant chooses to receive feedback when receiving feedback costs 10 sliders, and column (2) shows the other individuals' forecasted probability of choosing to receive feedback when receiving feedback costs 10 sliders. Women are significantly more likely than men to choose feedback by 4.71 pp ($p = 0.015$; Table 5, column (1)), but they are forecasted to be significantly less likely than men to choose feedback by 4.27 pp ($p < 0.001$; Table 5, column (2)). The forecasted gender difference of −4.27 pp lies outside the 95% confidence interval for the observed gender difference of (0.91 pp, 8.57 pp). In fact, we can reject any difference in the direction of men having greater demand for feedback than women.[29]

Table 5. Observed and Forecasted Probability of Choosing to Receive Feedback When It Costs 10 Sliders to Receive Feedback, with Both Studies Pooled

	Observed (1)	Forecasted (2)
Female	0.0471**	−0.0427****
	(0.0194)	(0.0121)
Cognitive Test Study	−0.0873****	0.0359***
	(0.0204)	(0.0121)
N	2,169	1,585
R^2	0.0146	0.0127

Notes. Column (1) pools observations from the cognitive test feedback study and the interview feedback study. Coefficient estimates from an OLS regression of the probability that the participant chooses to receive feedback when it costs 10 sliders to receive it, regressed on a female indicator and an indicator for the cognitive test study. Sample restricted to the 860 subjects with monotonic preferences for feedback in the cognitive test study and the 1,309 subjects with monotonic preferences for feedback over positive prices in the interview study. Column (2) pools observations from the cognitive test forecast study and the interview forecast study. Coefficient estimates from an OLS regression of the forecasted probability of choosing to receive feedback when it costs 10 sliders to receive it, regressed on an indicator that the forecast is about women's choices and an indicator for the cognitive test study. Sample restricted to the 744 forecasters in the cognitive test study with monotonic forecasts over nonzero prices for feedback and the 841 forecasters in the interview study with monotonic forecasts about preferences for feedback over positive prices, and it is always restricted to the forecast about the gender first elicited from the subject. Robust standard errors are in parentheses.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; ****$p < 0.001$.

Consistent with past work, we find in our two studies that participants' expectations of receiving good news are positively associated with their demand for feedback and that women are less confident in their performance than men, results that have been hypothesized in the literature to imply that women will shy away from performance feedback more so than men. In this way, our results may be somewhat of a surprise. In our setting, more confident people demand more feedback, and women are less confident on average. Despite this, women demand no less feedback than men on average. Together, this suggests that there are likely factors beyond confidence that predict demand for feedback and that, because of these other factors, controlling for confidence, women may even have a greater demand for feedback than men. Future work should investigate this further, unpacking the factors beyond confidence that predict demand for largely noninstrumental performance feedback.[30]

Investigations of this type may also help to further interpret some of the mixed findings in this literature. There are many dimensions that have been varied to some degree across studies, such as whether the feedback has instrumental or strategic value, whether feedback can be avoided entirely, whether the feedback is noisy, and what the costs of feedback are. It is worth noting that even given these substantial variations, most studies have found minimal evidence of gender differences in demand for feedback (Eil and Rao 2011, Castagnetti and Schmacker 2022, Möbius et al. 2022). The one study that does point to sizable gender gaps in demand for feedback is Sharma and Castagnetti (2023), where participants choose the informativeness of noisy feedback. Work that explores the relationship between demand for deterministic information and demand for informativeness could enhance our understanding of the behavioral forces that shape information acquisition decisions and shed further light on these across-study differences.

Future work should also consider additional settings, such as opportunities for face-to-face performance feedback, to understand whether there are indeed contextual factors that might generate a gender gap. This would also help us to better understand what factors predict demand for feedback and how these factors interact with gender. We consider only a few factors here, including overconfidence, difficulty of the task, and the possibility of bias in the feedback.

Promising avenues for future research are investigating more social factors, including whether the feedback is provided publicly or privately and the relationship between the person providing the feedback and the recipient. One could also consider whether gender differences in demand emerge when feedback must be sought more proactively. In our setting, individuals respond to an offer to receive or avoid feedback, yet in many educational and labor-market scenarios, the possibility of receiving feedback may not be as explicit. In these more ambiguous settings, might gender gaps in demand for feedback emerge?[31]

Individual motivations for receiving feedback also deserve more attention. Whereas our paper and others engage with ego management and curiosity, understanding how a desire to learn (or the need to signal a desire to learn) impacts demand for feedback would be valuable. Finally, considering the type of feedback would also be worthwhile: how do preferences vary depending upon whether the feedback is about relative versus absolute talents or whether the feedback is primarily evaluative as opposed to constructive?

We show that individuals expect there to be a gender gap in demand for feedback in our setting. With this, we take a first step into connecting our findings to the question of whether gender affects the supply of feedback. A person's willingness to give feedback may depend on whether that person thinks the other wishes to receive it; thus, inaccurate beliefs about preferences for feedback may prevent individuals from receiving the information they need to develop their skills and advance their careers.[32] Our study does not investigate the supply of feedback, including whether supply depends on beliefs about the preferences of the feedback recipient. Future work should explore whether and how the inaccurate beliefs we document impact the provision of feedback.

Importantly, our data do not allow us to speak to the question of whether men or women *should* demand more feedback: does more feedback make them better off? The answer to this question likely depends on a range of important factors, including how individuals update their beliefs about themselves in response to the feedback they receive and how valuable accurate beliefs about performance are in the context of interest. Our main contribution is instead to present evidence from multiple settings rejecting the hypothesis that women demand performance feedback less than men. We think this is a useful observation for researchers and practitioners looking to understand sources of gender gaps in self-confidence and career advancement. Future work should expand upon our studies to investigate the welfare implications of our findings, bringing together insights from the literature on overconfidence and belief updating to paint a fuller picture of how the supply of feedback, demand for feedback, and responses to feedback together shape gender gaps.

Endnotes

[1] Eil and Rao (2011) find some evidence that men have higher willingness to pay for the information than women among very confident participants and that women require a larger subsidy than men to receive the information among very underconfident participants, although these differences are not statistically significant.

[2] For reviews of the literature on belief-based utility, see, for example, Benabou and Tirole (2016) and Molnar and Loewenstein (2022).

[3] Köszegi (2006) and Weinberg (2009) present models that make the opposite prediction. In these models, agents derive utility from believing they have high ability; thus, low self-confidence agents are predicted to seek feedback on their ability in the hope of updating their beliefs upward, whereas high self-confidence agents are predicted to shun feedback to avoid the risk of revising their beliefs downward. Eliaz and Spiegler (2006) show that these models have difficulty explaining a positive relation between beliefs and demand for information. Others have explored preferences for information in domains that are not necessarily ego relevant and have similarly found that individuals seek information that is expected to deliver good news and avoid information that is expected to deliver bad news, even though avoiding it might be detrimental to decision making. Examples include investors looking up their portfolios more frequently in good market days (Karlsson et al. 2009, Sicherman et al. 2016) or individuals avoiding medical tests for fear of receiving a positive result (Thornton 2008, Oster et al. 2013, Ganguly and Tasoff 2017, Schünemann et al. 2023; for theoretical models, see Caplin and Eliaz 2003, Schweizer and Szech 2018; see also Klinowski and Paulsen 2013).

[4] We inform participants that this is a test of cognitive skills, but we do not mention the term ASVAB.

[5] Previous work that induces variation in confidence by manipulating task difficulty include Dargnies et al. (2019), Barron and Gravert (2022), and Möbius et al. (2022).

[6] See Danz et al. (2022) and Healy and Kagel (2022) for recent work on how incentives may affect belief reports.

[7] We see our measure of how sure participants feel about their guess as a measure of their certainty in their beliefs, essentially how tight or accurate is the prior belief, or what Moore and Healy (2008) refer to as *precision*.

[8] We bundle together feedback on absolute and relative performance on the cognitive test to simplify the experiment design and analysis and because we have no reason to expect that the factors that drive demand will differ dramatically for feedback about absolute and relative performance.

[9] A slider is completed by moving the indicator of a track bar to a target location in the range of integers from zero to 100. It takes roughly four to five seconds to complete a slider.

[10] This gender imbalance in the feedback study was not intentional and is unlikely to have been a product of men and women failing to complete the study at different rates because 96% of participants who pass the initial bot and human checks and start the study go on to complete the study.

[11] Throughout the paper, p-values come from two-sided t-tests except when noted otherwise.

[12] There is no significant treatment difference: 3.73 on the easy version and 3.87 on the hard version ($p = 0.334$).

[13] Assuming a normal distribution for the standard error, to obtain 80% power for a 95% confidence interval, the true effect size must be at least 2.8 standard errors away from zero (Gelman and Hill 2006).

Given our standard error for the female coefficient in column (1) of Table 1 equal to 0.033, we are powered to detect a male-female gap in demand for feedback of at least 9.24 pp. We can also use our estimates to identify the size of the male-female gap in demand for feedback that we can reject; using specification 3 from Table 1, we can reject men being three pp or more likely to demand feedback than women at the $p = 0.05$ level.

[14] We do not collect data on how long it takes participants to complete sliders. Murad et al. (2019) found that men perform better than women on the slider task under piece-rate compensation, suggesting that doing sliders might be less costly for men than for women. If this is true, we would be, if anything, underestimating the extent to which women are more willing to pay for costly feedback than men, given their higher cost for completing the same number of sliders.

[15] We also ask participants how certain they are about these beliefs on a scale from 1 to 5. The average degree of certainty about absolute (relative) performance is 2.61 (2.86). Women express significantly less certainty in their absolute and relative beliefs than men, conditional on true performance and measured optimism (Tables A.9 and A.10 in Online Appendix A).

[16] The ego management hypothesis would suggest that it is those individuals who are most certain about performing well who are most eager to receive performance feedback. We can investigate this with our data. Looking at beliefs of absolute performance, the effect of certainty on demand is directionally larger for participants whose believed score is equal to or larger than the median belief within treatment condition: the effect is 0.014 ($p = 0.504$; Table A.11, column (1) in Online Appendix A) for participants below the median belief and 0.031 ($p = 0.141$; Table A.11, column (2) in Online Appendix A) for participants above the median belief; however, in a pooled regression, the interaction between beliefs and certainty is not significant (Table A.11, column (3) in Online Appendix A). Similarly, for beliefs of relative performance, the effect of certainty is greater for participants whose believed rank is equal to or lower (i.e., better) than the median belief within treatment: the effect is 0.007 ($p = 0.746$; Table A.11, column (4) in Online Appendix A) for participants below the median belief and 0.039 ($p = 0.062$; Table A.11, column (5) in Online Appendix A) for participants above the median belief; however, in a pooled regression, the interaction between beliefs and certainty is not significant (Table A.11, column 6 in Online Appendix A).

[17] Importantly, the 2SLS approach rests on the assumption that treatment assignment affects the demand for feedback only through its effect on self-confidence. This exclusion restriction would be violated if, for example, subjects considered the test to be systematically more (or less) informative of their abilities in the hard version of the test, and this produced systematic differences in demand for feedback across treatments.

[18] The results are similar when controlling for the participants' reports of how influential and generalizable they find the feedback to be.

[19] Most subjects in the forecast study seem to have been confused about what it means to pay a cost to *avoid* feedback. In the online appendix, we discuss this issue in more detail and analyze forecasts of demand for feedback when it is costly to avoid feedback.

[20] Treatment, gender of the respondent, and order of the incentivized elicitation (forecast about men or women elicited first) are not significantly correlated with the answers to the qualitative Likert questions.

[21] We give subjects the opportunity to receive feedback only on their relative performance, not their absolute performance, because we felt that the interview score by itself was unlikely to convey much information on performance.

[22] We have some evidence that objective scores capture performance in the interview. For the 20 participants whose answers were rated by HR professionals, the correlation between objective scores and ratings assigned by the HR professional is 0.4 in the blind condition and 0.13 in the nonblind condition. Looking at the four facets that make up the objective scores, men score higher than women on intellectual curiosity and tolerance to stress, whereas women score higher on striving for achievement. There are no average gender differences in assertiveness, which is contrary to the stereotypical view that men are more assertive (Fiske et al. 2007), but is consistent with Coffman et al. (2021), who used text analysis of free-form conversation and found no gender differences in assertiveness as perceived by gender-blind coders, though the authors found a significant gender gap favoring men when coders were aware of conversant gender.

[23] A negative value for the maximum willingness to pay, $-X$, indicates that the subject is willing to complete up to X sliders to avoid receiving feedback.

[24] Given our standard error for the female coefficient in column (1) of Table 3 equal to 3.173, we are 80% powered to detect a male-female gap in maximum willingness to pay for feedback of at least 8.88 sliders.

[25] This is contrary to the cognitive test study, for which we found that performance on the test is negatively correlated with demand for feedback. Comparing performance measures across studies is difficult, though, because the correlation between AI score and demand for feedback on the interview may reflect an effect of performance but also personality (intellectual curiosity, a tendency to strive for achievement, assertiveness, and tolerance to stress) on the demand for feedback.

[26] Is this null effect because the treatment was not received? We have suggestive evidence that the treatment was received: looking at beliefs about average gender differences in the evaluations by HR professionals (a $1-5$ variable, where 5 is that HR professionals would rank men much better), in the blind condition, the mean belief is 3.01 for male participants versus 3.08 for female participants ($p = 0.226$). But when participants know that the HR evaluations will be nonblind, a gender gap emerges: the mean belief is 2.96 for men versus 3.21 for women ($p < 0.001$). In a regression, the difference in difference is significant ($p = 0.058$ without controls, $p = 0.091$ with controls). Thus, across subjects, we have some suggestive evidence that women seem to anticipate a modest amount of sex-based discrimination in evaluation.

[27] As in the cognitive test study, the positive relation between certainty of beliefs and the demand for feedback is concentrated among participants who are optimistic about their performance. The relation is much stronger and significant only for participants whose believed rank is equal to or lower (i.e., better) than the median belief within treatment (Table B.7 in Online Appendix B). This is another indication that individuals demand feedback in expectation of consuming good news.

[28] These results are similar when controlling for the participants' reports of how influential and generalizable they find the feedback to be.

[29] In the pooled analysis in Table 5, we are 80% powered to detect an observed gender difference of 5.4 pp and a forecasted gender difference of 3.4 pp, which are 7% of the mean observed uptake of feedback when it costs 10 sliders to receive it (73.7%) and 6% of the mean forecasted uptake (54.6%), respectively.

[30] More generally, our finding that overconfidence correlates with demand for feedback raises interesting questions for future work. When overconfident individuals receive feedback, they will be likely to receive (objectively) disappointing news on average. Viewed through the lens of long-run belief maintenance, this creates an apparent tension between remaining overconfident and continuing to demand more feedback. Zimmermann (2020) and others have begun to explore how individuals can maintain positive self-images in the face of negative feedback through motivated reasoning and biased memory. Future work should continue to investigate this tension, linking it to demand for feedback over time.

[31] This could be seen as analogous to results from the negotiation literature, which suggest smaller gender differences when it is clear that negotiation is a possibility compared with situations with greater ambiguity about whether it is appropriate to negotiate (Bowles et al. 2005).

[32] Closest to this question, Gallen and Wasserman (2022) show that students' genders affect the information they receive about careers, with individuals providing more information about work-life balance to women, and Dupas et al. (2021) find that female presenters receive more feedback during seminars. Understanding what beliefs—if any—underlie these differences is an important open question.

References

Alston M (2019) The (perceived) cost of being female: An experimental investigation of strategic responses to discrimination. Working paper, Department of Economics, Florida State University, Tallahassee, FL.

Araujo FA, Carbone E, Conell-Price L, Dunietz MW, Jaroszewicz A, Landsman R, Lame D, Vesterlund L, Wang SW, Wilson AJ (2016) The slider task: An example of restricted inference on incentive effects. *J. Econom. Sci. Assoc.* 2:1–12.

Ashford SJ, Blatt R, VandeWalle D (2003) Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *J. Management* 29(6):773–799.

Barron K, Gravert C (2022) Confidence and career choices: An experiment. *Scandinavian J. Econom.* 124(1):35–68.

Benabou R, Tirole J (2016) Mindful economics: The production, consumption, and value of beliefs. *J. Econom. Perspect.* 30(3):141–164.

Bordalo P, Coffman K, Gennaioli N, Shleifer A (2019) Beliefs about gender. *Amer. Econom. Rev.* 109(3):739–773.

Bowles HR, Babcock L, McGinn KL (2005) Constraints and triggers: Situational mechanics of gender in negotiation. *J. Personality Soc. Psych.* 89(6):951–965.

Brown R, Mansour H, O'Connell SD, Reeves J (2019) Gender differences in political career progression: Evidence from U.S. elections. Discussion Paper No. 12569, Institute of Labor Economics, University of Bonn, Bonn, Germany.

Burks SV, Carpenter JP, Goette L, Rustichini A (2013) Overconfidence and social signalling. *Rev. Econom. Stud.* 80(3):949–983.

Buser T, Yuan H (2019) Do women give up competing more easily? Evidence from the laboratory and the Dutch Math Olympiad. *Amer. Econom. J. Appl. Econom.* 11(3):225–252.

Buser T, Niederle M, Oosterbeek H (2014) Gender, competitiveness, and career choices. *Quart. J. Econom.* 129(3):1409–1447.

Caplin A, Eliaz K (2003) AIDS policy and psychology: A mechanism-design approach. *RAND J. Econom.* 34(4):631–646.

Castagnetti A, Schmacker R (2022) Protecting the ego: Motivated information selection and updating. *Eur. Econom. Rev.* 142: 104007.

Coffman K (2014) Evidence on self-stereotyping and the contribution of ideas. *Quart. J. Econom.* 129(4):1625–1660.

Coffman K, Klinowski D (2020a) A study of feedback aversion. Accessed January 2024, https://www.socialscienceregistry.org/trials/5990/history/70240.

Coffman K, Klinowski D (2020b) A study of feedback aversion on a job interview task. Accessed January 2024, https://www.socialscienceregistry.org/trials/6735/history/79497.

Coffman K, Klinowski D (2022) Beliefs about receiving feedback. Accessed January 2024, https://doi.org/10.1257/rct.10413-1.0.

Coffman K, Collis M, Kulkarni L (2023a) Stereotypes and belief updating. *J. Eur. Econom. Assoc.* 22(3):1011–1054.

Coffman K, Flikkema CB, Shurchkov O (2021) Gender stereotypes in deliberation and team decisions. *Games Econom. Behav.* 129:329–349.

Coffman K, Ugalde Araya P, Zafar B (2023b) A (dynamic) investigation of stereotypes, belief-updating, and behavior. NBER Working Paper No. 29382, National Bureau of Economic Research, Cambridge, MA.

Coutts A (2019) Good news and bad news are still news: Experimental evidence on belief updating. *Experiment. Econom.* 22(2):369–395.

Danz D, Vesterlund L, Wilson AJ (2022) Belief elicitation and behavioral incentive compatibility. *Amer. Econom. Rev.* 112(9):2851–2883.

Dargnies M-P, Hakimov R, Kübler D (2019) Self-confidence and unraveling in matching markets. *Management Sci.* 65(12):5603–5618.

De Quidt J, Vesterlund L, Wilson AJ (2019) Experimenter demand effects. Schram A, Ule A, eds. *The Handbook of Research Methods and Applications in Experimental Economics* (Edward Elgar Publishing, Cheltenham, UK).

Dupas P, Modestino AS, Niederle M, Wolfers J (2021) Gender and the dynamics of economics seminars. NBER Working Paper No. W28494, National Bureau of Economic Research, Cambridge, MA.

Eil D, Rao JM (2011) The good news-bad news effect: Asymmetric processing of objective information about yourself. *Am. Econom. J. Microeconom.* 3(2):114–138.

Eliaz K, Spiegler R (2006) Can anticipatory feelings explain anomalous choices of information sources? *Games Econom. Behav.* 56:87–104.

Ellison G, Swanson A (2023) Dynamics of the gender gap in high math achievement. *J. Human Resources* 58(5):1679–1711.

Exley CL, Kessler JB (2022) The gender gap in self-promotion. *Quart. J. Econom.* 137(3):1345–1381.

Exley CL, Nielsen K (2024) The gender gap in confidence: Expected but not accounted for. *Amer. Econom. Rev.* 114(3):851–885.

Exley CL, Fisman R, Kessler JB, Lepage LP, Li X, Low C, Shan X, Toma M, Zafar B (2024) Information-optional policies and the gender concealment gap. NBER Working Paper No. 32350, National Bureau of Economic Research, Cambridge, MA.

Fiske ST, Cuddy AJC, Peter G (2007) Universal dimensions of social cognition: Warmth and competence. *Trends Cognitive Sci.* 11(2): 77–83.

Fang C, Zhang E, Zhang J (2021) Do women give up competing more easily? Evidence from speedcubers. *Econom. Lett.* 205:109943.

Gallen Y, Wasserman M (2022) Informed choices: Gender gaps in career advice. Working paper, IZA - Institute of Labor Economics, Bonn, Germany.

Ganguly A, Tasoff J (2017) Fantasy and dread: The demand for information and the consumption utility of the future. *Management Sci.* 63(12):4037–4060.

Gelman A, Hill J (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, New York).

Gill D, Prowse V (2012) A structural analysis of disappointment aversion in a real effort competition. *Amer. Econom. Rev.* 102(1):469–503.

Gill D, Prowse V (2014) Gender differences and dynamics in competition: The role of luck. *Quant. Econ.* 5(2):351–376.

Golman R, Hagmann D, Loewenstein G (2017) Information avoidance. *J. Econom. Lit.* 55(1):96–135.

Golman R, Loewenstein G, Molnar A, Saccardo S (2022) The demand for, and avoidance of, information. *Management Sci.* 68(9):6454–6476.

Grosse ND, Reiner G (2010) Explaining gender gaps in competitiveness: Gender-task stereotypes. Jena Economic Research Papers No. 2010017, Friedrich Schiller University, Max Planck Institute of Economics, Jena, Germany.

Healy PJ, Kagel J (2022) Testing elicitation mechanisms via team chat. Working paper, The Ohio State University, Columbus, OH.

Kang L, Lei Z, Song Y, Zhang P (2024) Gender differences in reactions to failure in high-stakes competition: Evidence from the national college entrance exam retakes. *J. Political Econom. Microeconomics* 2(2):355–397.

Karlsson N, Loewenstein G, Seppi D (2009) The ostrich effect: Selective attention to information. *J. Risk Uncertainty* 38(2):95–115.

Klinowski D (2019) Selection into self-improvement and competition pay: Gender, stereotypes, and earnings volatility. *J. Econom. Behav. Organ.* 158:128–146.

Klinowski D, Paulsen JS (2013) Impaired awareness in Huntington disease: Medical evidence in relation to the optimal expectations model. Working paper, University of Pittsburgh, Pittsburgh, PA.

Köszegi B (2006) Ego utility, overconfidence, and task choice. *J. Eur. Econom. Assoc.* 4(4):673–707.

Masatlioglu Y, Orhun AY, Raymond C (2023) Intrinsic information preferences and skewness. *Amer. Econom. Rev.* 113(10):2615–2644.

Möbius MM, Niederle M, Niehaus P, Rosenblat TS (2022) Managing self-confidence: Theory and experimental evidence. *Management Sci.* 68(11):7793–7817.

Molnar A, Loewenstein G (2022) Thoughts and players: An introduction to old and new economic perspectives on beliefs. Musolino J, Sommer J, Hemmer P, eds. *The Cognitive Science of Belief: A Multidisciplinary Approach* (Cambridge University Press, Cambridge, UK), 321–350.

Moore DA, Healy PJ (2008) The trouble with overconfidence. *Psych. Rev.* 115(2):502–517.

Murad Z, Stavropoulou C, Cookson G (2019) Incentives and gender in a multi-task setting: An experimental study with real-effort tasks. *PLoS One* 14(3):e0213080.

Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* 122(3):1067–1101.

Oster E, Shoulson I, Dorsey ER (2013) Optimal expectations and limited medical testing: Evidence from Huntington disease. *Amer. Econom. Rev.* 103(2):804–830.

Pereda P, Montoya Diaz MD, Rocha F, Matsunaga L, Borges BP, Mena-Chalco J, Narita R, Brenck C (2023) Are women less persistent? Evidence from submissions to a nationwide meeting of economics. *Appl. Econom.* 55(16):1757–1768.

Schünemann J, Strulik H, Trimborn T (2023) Anticipation of deteriorating health and information avoidance. *J. Health Econom.* 89:102755.

Schweizer N, Szech N (2018) Optimal revelation of life-changing information. *Management Sci.* 64(11):5250–5262.

Sharma K, Castagnetti A (2023) Demand for information by gender: An experimental study. *J. Econom. Behav. Organ.* 207:172–202.

Shastry GK, Shurchkov O, Xia LL (2020) Luck or skill: How women and men react to noisy feedback. *J. Behav. Experiment. Econom.* 88:101592.

Shurchkov O (2012) Under pressure: Gender differences in output quality and quantity under competition and time constraints. *J. Eur. Econom. Assoc.* 10(5):1189–1213.

Sicherman N, Loewenstein G, Seppi DJ, Utkus SP (2016) Financial attention. *Rev. Financial Stud.* 29(4):863–897.

Stone D, Heen S (2014) *Thanks for the Feedback: The Science and Art of Receiving Feedback Well* (Penguin Group, New York).

Thornton RL (2008) The demand for, and impact of, learning HIV status. *Amer. Econom. Rev.* 98(5):1829–1863.

Wasserman M (2023) Gender differences in politician persistence. *Rev. Econom. Statist.* 105(2):275–291.

Weinberg BA (2009) A model of overconfidence. *Pacific Econom. Rev.* 14(4):502–515.

Zimmermann F (2020) The dynamics of motivated beliefs. *Amer. Econom. Rev.* 110(2):337–363.