

Women demand performance feedback no less than men

Katherine Coffman ^a and David Klinowski ^b

December 2022

Abstract

Across two experiments, 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 job interview, whether the feedback is objective or subjective, and whether 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 they perceive feedback to be more influential and generalizable for self-assessment than men do. 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. Our results suggest that inaccurate beliefs about women's demand for feedback may pose a more significant barrier to development and advancement than a gender gap in preferences for feedback.

^a Harvard Business School. kcoffman@hbs.edu.

^b Katz Graduate School of Business, University of Pittsburgh. dklinowski@katz.pitt.edu.

1 Introduction

Performance feedback plays a critical role in economic decision-making. 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 informs decisions about what educational tracks and career paths to pursue. Of course, in many important contexts, opportunities for feedback are not exogenously assigned: individuals have the ability to seek out, or avoid, performance feedback. For instance, an individual could proactively request feedback from a wide set of mentors, or avoid going to office hours. These decisions about whether, when, and how intensively to pursue performance feedback may have important implications for the quantity and quality of information an individual has about her 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. 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, Blatt, and Vandewalle, 2003) might in 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 is critical for understanding economic outcomes.

In this paper, we explore preferences for feedback, focusing on gender differences. If men and women vary in their preferences for feedback, this could lead to gaps in actions like coming forward with early-stage ideas, initiating career conversations with a manager, or even applying

for a new opportunity, potentially contributing to gender gaps in labor market outcomes. Our goal is to cleanly identify any gender difference in demand for performance feedback. This is hard to accomplish 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 of 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.

Previous literature has focused on two central factors in predicting demand for performance feedback: first, how useful will that information be (the instrumental channel), and second, how much consumption utility will the information provide (the consumption channel). In principle, either factor could contribute to a gender difference in demand. From an instrumental standpoint, it could be that there is a difference in how useful men and women believe performance feedback will be -- for instance, differences in beliefs of the quality or generalizability of the feedback, or differences in how confident individuals are in the information they already have. From a consumption standpoint, there could be gender differences in the consumption utility of receiving performance feedback; one might think of this as differences in how “thick-skinned” individuals are, or how much they enjoy hearing about their success. Our studies are designed to speak to each of these channels, asking how each matters for understanding preferences for feedback by gender.

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 different from, and sometimes greater 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 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 we hire and based upon 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' beliefs about their own performance, including both beliefs of how they performed and how certain they are about those beliefs. This allows us to examine an important aspect of the consumption channel, asking whether people who expect better performance have greater demand for feedback. To help us probe the instrumental channel, we

elicit participants' self-reports of how much they expect the feedback to influence their own personal evaluation of (i) their performance on the task, (ii) their abilities in this domain more generally, and (iii) their capabilities in other aspects of life (an elicitation that has been used in psychology, e.g., Roberts and Nolen-Hoeksema, 1989).

Across our two studies, we find evidence that both consumption and instrumental motivations contribute to the demand for feedback, consistent with past work that we review in the next section. Participants who have more optimistic beliefs of their performance, and who are more sure of their beliefs, have greater demand for feedback, which suggests that subjects seek feedback in expectation of consuming good news. Similarly, subjects who perceive that the feedback will have a greater influence on their self-assessments have greater demand for feedback, which suggests that subjects seek feedback for its evaluative, or instrumental, value. While we see some evidence of gender differences in beliefs about own performance (men more optimistic than women) and in beliefs about the instrumental value of the feedback (women placing higher value on it than men), on net women are no more feedback averse than men, and in some specifications women's demand for feedback is significantly greater than men's.

We ask individuals from a separate population to predict our study participants' demand for feedback. We collect incentivized beliefs of the share of male (female) participants that opt for feedback. Figure 1A 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 opt for feedback across our two studies (t-test $p < 0.001$), and beliefs about men stochastically dominate beliefs about women ($p < 0.001$ 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 greater demand for feedback is approximately

twice as common as the belief that women have 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, while we see that women in our studies demand performance feedback no less than men, others believe them to be more feedback averse than men. These inaccurate beliefs could drive under-provision of feedback to women in these economically-important contexts.

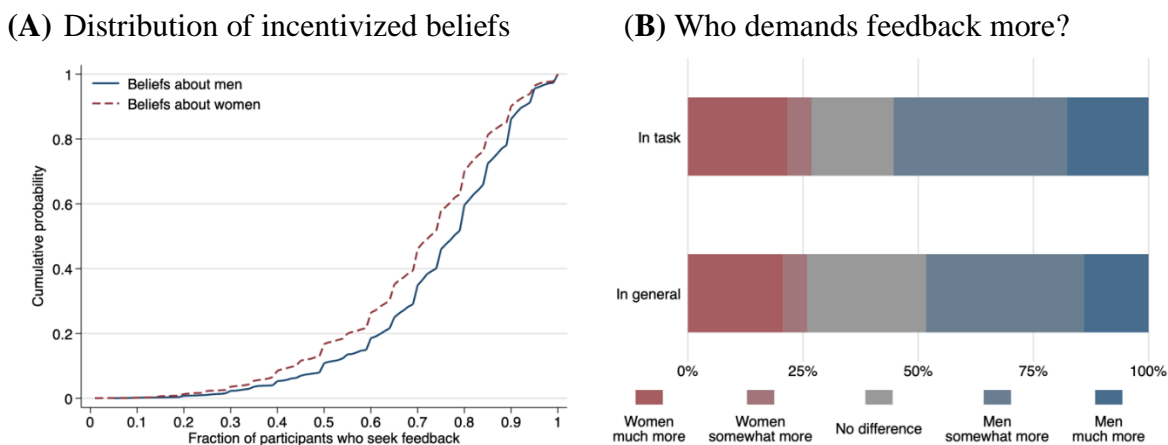


Figure 1 Beliefs of others' demand for feedback

Notes: Panel A plots the cumulative distribution of 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. Panel B shows responses to the 1-5 Likert questions "Overall, for the task participants completed in the previous stud, how would you describe differences in men's and women's preferences for finding out how they performed?" (In task), and "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?" (In general). Figures pool observations from the Cognitive Test and Interview studies.

2 Related Literature

A recent literature explores theoretically and empirically motivations that might affect a person's demand for feedback, or demand for information more generally, beyond improving decision-making. One primary alternative motivation is ego-management. For instance, in belief-based anticipatory utility models (Kőszegi, 2006; Weinberg, 2009), agents derive utility from

believing they have high ability, and so low self-confidence agents are predicted to seek feedback on their ability in the hope of updating their beliefs upward, while high self-confidence agents are predicted to shun feedback to avoid the risk of revising their beliefs downward.¹ Attention-based anticipatory utility models (Karlsson, Loewenstein, and Seppi, 2009; Ganguly and Tasoff, 2017; Golman and Loewenstein, 2018; Bolte and Raymond, 2022) make the opposite prediction. In these latter models, agents derive utility directly from the information they consume, and so high-confidence agents are predicted to be more likely to seek feedback on their ability. Empirical evidence largely supports the ego-management motivation, particularly the information-consumption mechanism (Karlsson, Loewenstein, and Seppi 2009, Burks et al. 2013, Eil and Rao 2011, and Masatlioglu, Orhun, and Raymond 2021). Our results are also consistent this motivation, as we find across our two studies that subjects with more optimistic beliefs of their performance have greater demand for performance feedback.

The role of gender in the demand for feedback is less explored in the literature. All else fixed, the attention-based anticipatory utility motivation suggests that higher self-confidence among men would lead them to be more feedback-seeking relative to women.² And, work in psychology and economics finds that women tend to regard evaluative feedback as more informative about abilities and self-worth than men do, particularly with feedback that is subjective or noisy (Roberts and Nolen-Hoeksema, 1989; Dweck, 2000; Shastry, Shurchkov, and Xia, 2020). Recent studies provide some suggestive empirical evidence. In their Supplementary Materials, Mobius et al. (2022) document that men and women have similar average valuations for feedback

¹ Eliaz and Spiegler (2006) that show that belief-based anticipatory utility models have difficulty explaining a positive relation between beliefs and the demand for information.

² Previous work also finds women to be less confident in their performance on tasks than men, especially on male-typed tasks (e.g., Niederle and Vesterlund, 2007; Grosse and Reiner, 2010; Shurchkov, 2012; Coffman, 2014; Buser, Niederle, and Oosterbeek, 2014; Bordalo et al., 2019; Klinowski, 2019).

on relative performance on a cognitive test, but that women are more likely to be willing to pay to avoid receiving feedback. Eil and Rao (2011) uncover that, among very confident participants, men have directionally higher demand for feedback on their cognitive ability and physical attractiveness than women and that, among very underconfident participants, women require a larger subsidy than men to receive feedback.

We contribute to this literature by examining gender differences in demand for feedback across two well-powered, pre-registered studies specifically designed to address gender. Our data can not only measure a potential gap but also speak to its sources. We couple this evidence with additional data on beliefs about gender differences in demand for performance feedback, allowing us to contrast true differences with anticipated differences.

3 Demand for Feedback on a Cognitive Test

3.1 Experimental Design: Preferences for Feedback Study

In the Preferences for 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 5 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).³ We intentionally chose questions from stereotypically male-typed domains. This increases the extent to which our results are likely to speak to real world

³ We inform participants that this is a test of cognitive skills, but do not mention the term ASVAB.

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.

We randomize participants into either an easy or a hard version of the test, in a between-subjects design. The reason for these treatments is to generate exogenous variation in performance beliefs, and, in doing so, increase our ability to causally explore the role of beliefs in explaining the demand for feedback. This rationale builds on previous findings that beliefs about both absolute and relative ability depend upon the difficulty of the task (Moore and Healy, 2008; Bordalo et al., 2019). 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.⁴ Participants then indicate how they think they ranked relative to 9 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.

⁴ See Danz, Vesterlund, and Wilson (2020) and Healy and Kagel (2022) for recent work on how incentives may affect belief reports.

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 9 participants (as a single bundle of information), on a 1-5 scale ranging from "Not at all interested" to "Extremely interested". This is an unincentivized report, since 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 9 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 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).⁵ 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

⁵ A slider is completed by moving the indicator of a track bar to a target location in the range of integers from 0 to 100. It takes roughly 4 to 5 seconds to complete a slider.

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 US. 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 3 more questions correctly than men, (ii) women answered 1 or 2 more questions correctly than men, (iii) women and men answered correctly the same number of questions, (iv) men answered 1 or 2 more questions correctly than women, and (v) men answered at least 3 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 1-10 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.

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. In the instructions to the elicitation of preferences over feedback, we inform participants that they will be required to type in their feedback if they receive it. 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 of them 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 in 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 US, with at least 100 previous HITs completed on MTurk, and with approval rating of at least

95%. Participants had to pass several comprehension and attention checks distributed throughout the session to complete the study. We pre-registered the study before data collection (Coffman and Klinowski, 2020a).

3.2 *Experimental Design: Beliefs Study*

After conducting the Preferences for Feedback Study, we elicited beliefs about participants' demand for feedback from a separate sample of subjects. We call this the Beliefs Study. We first elicit their demographic information (gender, age bracket, and region of residence), and then inform them about the Preferences for 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. Then, to guess how many out of 100 [male/female] participants chose to receive feedback when it cost 10 sliders to receive feedback. Finally, 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 question always in the same order within gender. Subjects receive a bonus of \$0.25 if one of their six guesses, randomly chosen, is within 5 percentage points 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 US, with at least 100 completed studies on the platform, and with approval rate of at least 95%. We pre-registered 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 on the Interview, as described in Section 4.

3.3 Results

3.3.1 Descriptive Statistics

The sample consists of 350 women and 645 men in the Preferences for Feedback Study, and 463 women and 489 men in the Beliefs Study.⁶ Table A1 in Appendix A provides descriptive statistics, and Table A2 shows that demographic characteristics are largely balanced across treatments (easy and hard version of the test) within study. As expected, participants answer significantly more questions correctly on the easy version than the hard version of the test (9.4 vs. 7.2, $p < 0.001$). There are no gender differences in test scores in either treatment (Table A1).

On average, participants overestimate their scores. Average beliefs of absolute score are 11.4 on the easy version and 9.6 on the hard version ($p < 0.001$). Panels A-B of Figure A1 plot beliefs about test score as a function of test score. Conditional on test performance, women's beliefs are on average less optimistic than men's on both versions of the test (columns 1-2 of Table A3). For any given test score, beliefs tend to be more pessimistic in the hard version (column 3 of Table A3), indicating that, as expected, our treatment assignment produces exogenous variation in beliefs. This treatment effect on beliefs is not significantly different across gender (n.s. interaction term in column 4 of Table A3).

In terms of beliefs about 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 ($p < 0.001$), where rank of 1 is best and 10 is worst. On average, women are significantly less optimistic than men about their relative performance (panels C-D of Figure A1). Conditional on true rank, women's beliefs of rank are approximately 0.8 ranks lower than men's ($p < 0.01$, column 3 of Table A4). Both the easy and the hard test produce this gender gap (columns 1-2 of Table

⁶ This gender imbalance in the Preferences for Feedback Study was not intentional.

A4), with no significant difference in the gap across treatments (n.s. interaction term in column 4 of Table A4). 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).⁷ Women express significantly less certainty in their absolute and relative beliefs than men, conditional on true performance and measured optimism.

3.3.2 *Preferences over Feedback*

We present results in this section for subjects with monotonic preferences for feedback, which constitute 86.4 percent of the sample. Failure of monotonicity does not vary significantly across treatment ($p=0.869$) or gender ($p=0.287$). Table A7 shows that failure of monotonicity is positively correlated with finding the instructions difficult and with lower test score, suggesting that non-monotonicity is most likely related to confusion.

Overall, we find substantial demand for feedback. Looking first at the qualitative measure asking how interested subjects were in learning about how they performed, the average is 3.78 on the 1 – 5 scale.⁸ We find that women report weakly 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 while women report on average 4.03 ($p=0.004$). On the hard version, men report on average 3.75 while 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 A2 plots the cumulative probability of the qualitative measure of interest in receiving feedback across treatment and gender.

⁷ Interestingly, there are clear, strong correlates of expressed certainty about beliefs (Tables A5-A6). First, individuals who perform worse on the test express more certainty about their beliefs. Second, *conditional on measured performance*, individuals who are more optimistic about their performance also express more certainty about their beliefs (both for absolute and relative measures of beliefs). Third, individuals assigned to the hard version of the test express less certainty about their beliefs, even conditional on true performance and measured optimism.

⁸ There is no treatment difference: 3.73 on the easy version, and 3.87 on the hard version ($p=0.334$).

Figure 2 presents the fraction of subjects who choose to receive feedback for each of the three price-list items and two treatment conditions, separated by gender. We see that 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 falls as it becomes more costly to receive ($p < 0.001$ for all pairwise comparisons), suggesting attentiveness and understanding among these participants. More critically, there is no significant gender gap in demand for feedback at any price in either treatment.

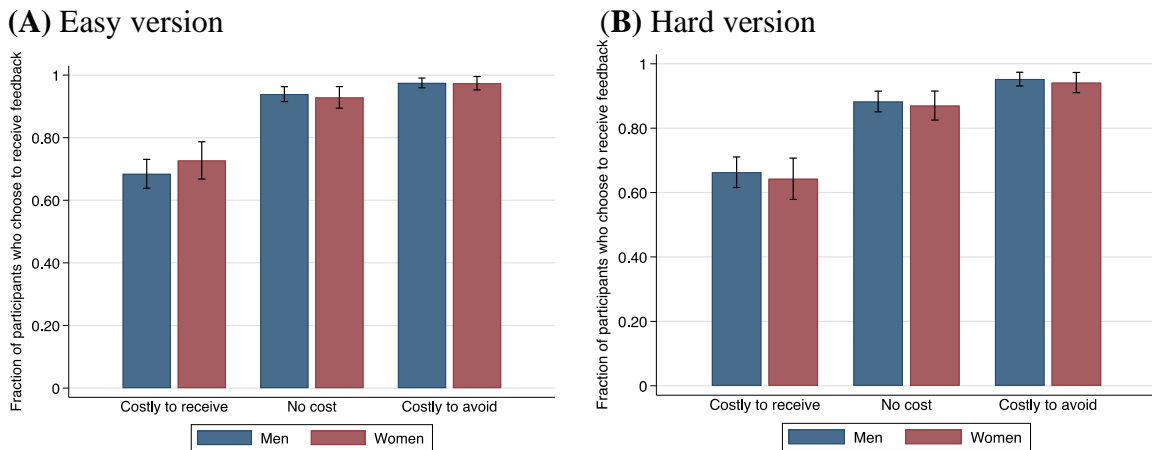


Figure 2. Demand for feedback on cognitive test performance

Notes: Observations from the Cognitive Test Preferences for Feedback Study. Sample restricted to subjects with monotonic preferences for feedback over all prices. Whiskers indicate 90-percent confidence intervals.

Next, we examine what factors help to explain demand, and we ask whether conditioning on these factors changes the (null) gender gap. We focus on the choice of receiving feedback when it is costly to do so, since 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 predicts the probability of choosing costly feedback. Specifications 1-2 show that demand is not significantly different across treatment or gender, irrespective of the

inclusion of demographic controls. We estimate that the 95% confidence interval of the gender gap (women relative to men) in the likelihood of choosing costly feedback is [-0.038, 0.092].

Specification 3 shows that there is no interaction effect between treatment and gender.

Table 1. Demand for costly feedback on cognitive test performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.011 (0.033)	0.027 (0.033)	0.070 (0.046)	0.031 (0.033)	0.041 (0.033)	0.053 (0.034)	0.058* (0.033)	0.061* (0.033)
Hard	-0.044 (0.032)	-0.048 (0.032)	-0.017 (0.040)	-0.048 (0.032)	-0.033 (0.032)	-0.028 (0.032)	-0.020 (0.032)	-0.016 (0.031)
Female x Hard			-0.087 (0.066)					
Actual rank (1:best , 10:worst)				0.014*** (0.006)	0.016*** (0.006)	0.015*** (0.006)	0.009* (0.006)	0.009* (0.006)
Belief of test score					0.005** (0.002)			
Certainty in test sc.					0.023 (0.014)			
Belief of rank (1:best , 10:worst)						-0.023**** (0.006)	-0.017*** (0.006)	-0.019*** (0.006)
Certainty in rank						0.029** (0.014)	0.011 (0.015)	0.010 (0.015)
Positive influence							0.081**** (0.018)	
Negative influence								0.089**** (0.018)
Controls	N	Y	Y	Y	Y	Y	Y	Y
N	860	860	860	860	860	860	860	860
R ²	0.0024	0.0262	0.0281	0.0338	0.0448	0.0541	0.0792	0.0836

Notes: Observations from the Cognitive Test Preferences for Feedback Study. Coefficient estimates from OLS regressions of the probability that the participant chooses to receive feedback when it is costly to receive it. *Positive influence* (*Negative influence*) is the first component derived from principal component analysis of the three questions on the perceived influence of better (worse) than expected feedback. Controls are age, race, region of residence, high school in the US, and the order of the three questions that elicit preferences for feedback. Columns 7-8 control in addition for the order of the questions on the influence of receiving better or worse than expected feedback. Sample restricted to subjects with monotonic preferences for feedback. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

Performing worse, having more optimistic beliefs, and being more certain of those beliefs are each correlated with greater demand for feedback.⁹ Table 1, Specification 4 shows that 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 treatment condition. Specification 5 includes beliefs about absolute performance and certainty of these beliefs, and specification 6 includes instead beliefs about relative performance and certainty of these beliefs. In both specifications, conditional on actual rank, more optimistic beliefs are associated with greater demand for feedback.¹⁰ Greater certainty about these beliefs is associated with greater demand for feedback (directionally so for beliefs about absolute performance, and statistically significantly so for beliefs about relative performance).¹¹ In neither specification is gender significant. These results are similar when the outcome is the choice of receiving feedback when there is no cost (Table A9) and when it is costly to avoid feedback (Table A10).

These results seem inconsistent with a story where demand for information is primarily instrumental. We would expect feedback to be most useful in cases where participants were least

⁹ For the unincentivized measure of interest in receiving feedback, more optimistic beliefs, particularly of relative performance, and greater certainty about one's beliefs predict greater interest in feedback (Table A11). Note that for this measure, we find that women are significantly more interested than men in receiving feedback, particularly on the easy version of the test.

¹⁰ Table A13 examines whether self-confidence causally affects the demand for feedback by using random assignment to the easy and hard versions of the test to instrument for beliefs of relative performance. Estimates of the effect of beliefs on the demand for feedback using the instrumental variable approach (2SLS) are more than twice as large as the OLS estimates (-0.049 vs. -0.018 in specification (1), and -0.043 vs. -0.019 in specification (2)), suggesting that beliefs causally affect feedback preferences. However, as is often the case, the 2SLS estimates are also much less precise than the OLS estimates, to the point of becoming insignificant ($p=0.335$ and $p=0.368$ in specifications (1) and (2), respectively). Note that the instrumental variables approach rests on the assumption that treatment assignment affects the demand for feedback only through its effect on confidence. This exclusion restriction would be violated if, for example, subjects in the hard version of the test systematically thought that the test was more, or less, informative than the easy version of the test and that led them to systematic differences in demand for feedback across treatments.

¹¹ However, the interaction between beliefs and certainty is not significant, indicating that the positive association between demand for feedback and optimism does not depend on how certain subjects are of their beliefs (see Table A8, which replicates specifications 5 and 6 of Table 1, but now including as additional control the interaction between beliefs of absolute or relative ability and certainty of these beliefs).

sure about their beliefs. Instead, we find that those who are most sure about their beliefs are significantly more willing to pay for feedback. Our results seem most consistent with an explanation where individuals who are most sure that they will receive “good news” are most likely to opt-in to receiving feedback. To probe this further, we can consider how self-reports of how influential feedback would be on their beliefs predict their demand for feedback.

Overall, individuals forecast a non-trivial impact of feedback on their beliefs of their own abilities. Using the 1-10 scale, with 10 being extremely impactful, on average participants assess the impact of feedback will be 5.85 on their own evaluation of their performance, 5.72 on their cognitive ability generally, and 5.15 on their capabilities in other aspects of life. On average, participants report significantly greater expected impact of better than expected feedback than worse than expected feedback.¹² Table A12 shows that women report somewhat more influence of the feedback on their evaluation of their performance on this specific task ($p=0.078$ for better than expected feedback, n.s. for worse than expected feedback). But, when thinking about the impact of feedback on their views of their cognitive ability generally or their capabilities in other aspects of life, women expect, if anything, less impact of feedback (significantly so when assessing other aspects of life, $p=0.044$). We also observe that individuals who perform worse on the test and individuals with more optimistic beliefs about their performance on the test expect the feedback to be more influential.

Columns 7-8 in Table 1 augment column 6 of Table 1 by including self-reports of how participants interpret and generalize feedback. We do so by adding as a regressor the first component derived from a principal component analysis (PCA) of the influence questions. We

¹² Comparing better than expected to worse than expected feedback, we find that average reports are, respectively, 6.16 vs. 5.57 for own evaluation of performance, 5.97 vs. 5.48 for cognitive ability, and 5.36 vs. 4.94 for capabilities in other aspects of life, with $p<0.001$ for all three comparisons.

conduct the PCA separately for the three variables that involve better-than-expected feedback and the three variables that involve worse-than-expected feedback. For both sets of three variables, the high correlation between any two of the three variables—correlations ranging from 0.62 to 0.70—implies that the first principal component alone explains approximately 80 percent of the variation in the variables. We find that greater self-reported influence of feedback is significantly associated with greater demand. Moreover, conditional on performance, beliefs, and reported influence of feedback, women are *more* likely than men to demand costly feedback. The results in Table 1 are similar if we add as regressors the 6 influence questions separately rather than the first principal components only.

Thus, the results suggest both instrumental and consumption channels contribute to the demand for feedback: participants who report a greater expected impact of the feedback have significantly greater demand, as do participants with more optimistic beliefs who might anticipate receiving “good news.” But even after accounting for both channels with the variables we elicit in the study, we find no evidence that women are more feedback averse than men.

3.3.3 Beliefs of Preferences over Feedback

In this section we examine the Beliefs Study data, beliefs about demand for feedback on the cognitive test. Mirroring the analysis presented in the previous section, we compare the believed fraction of men and women who would demand costly feedback to the realized fraction of men and women who demanded it. We restrict the sample to the 76% of participants who have monotonic beliefs over nonzero prices (participants who believed that more subjects opt to receive feedback when it is free than when it is costly). In the Appendix, we replicate the analysis for

beliefs of the choice of receiving feedback at zero price, and for the full sample of participants rather than just participants with monotonic beliefs, obtaining similar results (Tables A16-A17).¹³

Table 2 estimates the guessed 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 a percentage point, to allow for easier comparability with our results from the demand for feedback study. Specification 1 performs a within-subject analysis, regressing the believed probability of demanding feedback on an indicator that the belief is about women and subject fixed effects. We estimate that participants guess 58.5% of men and 55.0% of women opt for feedback, a difference of 3.5 percentage points ($p < 0.001$).

Table 2. Beliefs of demand for costly feedback on the Cognitive Test

	(1)	(2)	(3)
Female	-0.035**** (0.007)	-0.044** (0.017)	-0.044** (0.017)
Hard		-0.027 (0.017)	-0.027 (0.017)
Female self			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 Beliefs Study. Coefficient estimates from OLS regressions of the guessed probability of opting for feedback when it is costly. Controls are age bracket and region of residence. Sample restricted to subjects with monotonic beliefs about preferences for feedback with nonzero prices for feedback. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$.

¹³ Most subjects in the Beliefs Study seem to have been confused about what it means to pay a cost to *avoid* feedback. In the Appendix, we discuss this issue in more detail and analyze beliefs of demand for feedback when it is costly to avoid feedback.

Specification 2 performs an across-subjects analysis, restricting the sample to only the first beliefs provided by the participant (recall that subjects are randomized into being asked first about men or first about women) and controlling for demographics (the belief provider's age bracket and region of residence). We estimate a larger believed gender gap in this specification, with participants believing that women are 4.4 percentage points less likely to demand feedback ($p=0.011$). Specification 3 shows that controlling for the subject's gender does not eliminate the believed gender gap. In Appendix Table A15, we show that the believed gender gap in demand for feedback is indistinguishable across the hard and easy versions of the cognitive test, and that both men and women believe that women have lower demand for feedback.

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 0 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 (t-test $p<0.001$) for the cognitive test question and 0.28 (t-test $p<0.001$) for the general question.¹⁴

In sum, we find that demand for feedback on performance on a cognitive test is high, and women are no less eager for this feedback than men. Despite this, others anticipate that women demand feedback less than men. In this study, 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

¹⁴ Table A18 shows that treatment, gender of the respondent, and order of the incentivized elicitation (beliefs about men or women elicited first) do not affect the responses to the unincentivized questions.

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

4.1 Experimental Design: Preferences for Feedback Study

The Preferences for 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. While we are interested in the sex of the participant, we collect also 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, Vesterlund, and Wilson, 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, if they submit an answer in two minutes or less, or if they 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 a HR professional, and we give them an opportunity to receive feedback on how they ranked on the evaluation relative to other participants.

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 a 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.

Prior Beliefs: After participants receive the preliminary information, we elicit their beliefs about their rank on the Interview Score relative to the other 9 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.

Preferences over Feedback: We elicit participants' preferences for receiving feedback on their relative performance on the interview in two parts.¹⁵ 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.

¹⁵ We give subjects the opportunity to receive feedback only on their relative performance, not their absolute performance, since we felt that the Interview Score by itself was unlikely to convey much information on performance.

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 (A) being told the rank on the Interview Score or (B) not being told the rank on the Interview Score. This question is followed by a block of 5 questions that involve a choice between (A) being told the rank on the Interview Score and completing X sliders, or (B) not being told the rank on the Interview Score. This block is followed in turn by a second block of 5 questions that involve a choice between (A) being told the rank on the Interview Score, or (B) not being told the rank on the Interview Score and completing X sliders. Within a block of 5 questions, X always increases from 2, 5, 10, 50, to 100. (We estimate that it would take participants 7-8 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 5-question blocks is randomized.

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 US. 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.

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 Study 1, 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, 3 weeks apart from each other, and 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 US, with at least 500 previous HITs completed on MTurk, and with approval rating 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 pre-registered 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 2 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 a HR professional. Note that Session 1 is identical across treatment; 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 pre-registered the Interview Preferences for Feedback Study (see Coffman and Klinowski, 2020b).

4.2 *Experimental Design: Beliefs Study*

After conducting the Preferences for Feedback Study, we elicited beliefs about the participants' demand for feedback, from a separate sample of subjects. We call this the Beliefs Study. From these new subjects, we first elicit their demographic information (gender, age bracket, and region of residence), and then inform them about the Preferences for 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' beliefs of the previous participants' demand for feedback. We elicit beliefs 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 subject. Subjects receive a bonus of \$0.25 if one randomly-selected guess (of the 22 total guesses, 11 for each gender) is within 5 percentage points of the correct answer.

Finally, we elicit beliefs about gender differences in demand for feedback using two, unincentivized qualitative questions, one asking about beliefs about the gender gap in demand for performance feedback on the interview task specifically and the other asking about 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 beliefs about demand for performance feedback on the cognitive test (see Section 3.2), 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

4.3.1 Descriptive Statistics

Table B1 in Appendix B provides descriptive statistics. The sample for the Preferences for Feedback Study consists of subjects that participated in Session 2, which are the participants for whom we collected the full set of demographic information and for whom we elicited their demand for feedback. Approximately 89 percent of Session-1 subjects completed Session 2. Attrition from Session 1 to Session 2 was directionally smaller in the blind condition than the nonblind condition (10 percent vs. 12 percent, chi-squared test $p=0.153$), and directionally smaller for women than for men (10 percent vs. 12 percent, 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 randomness. Moreover, since 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 a 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 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.¹⁶ Similar to the Cognitive Test study, beliefs of relative performance are close to the middle of the pack on average, with a mean believed rank of 4.6 in the blind treatment and 4.7 in the non-blind treatment ($p=0.442$).¹⁷ Women's beliefs are on average more pessimistic than men's conditional on AI score rank (Table B3). In terms of how certain participants are of their beliefs, the average degree of certainty is 2.54 on the 1 – 5 scale, with no significant gender difference (Table B4).

¹⁶ We have some evidence that objective scores capture performance on 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. Across all subjects in Session 2, higher objective scores predict more optimistic beliefs of relative performance ($p\text{-val}=0.030$ from an OLS regression of beliefs on objective scores with no controls), although this correlation becomes insignificant after controlling for demographics (Table B3). Looking at the four facets that make up the objective scores, men score higher than women on intellectual curiosity and tolerance to stress, while 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, Cuddy, and Glick, 2007), but is consistent with Coffman, Flikkema, and Shurchkov (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.

¹⁷ Figure B1 plots rank beliefs as a function of AI score rank.

4.3.2 Preferences over Feedback

We present results in this section for subjects with monotonic preferences for feedback, which constitute 92 percent of Session 2 participants in the Preferences for Feedback Study.¹⁸

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 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 (t -test $p=0.615$ from an OLS regression with no controls). To illustrate these results, Figure B2 plots the cumulative probability of the qualitative measure of interest in receiving feedback across treatment and gender. Table B6 shows similar results from regressions after controlling for demographics and other variables.

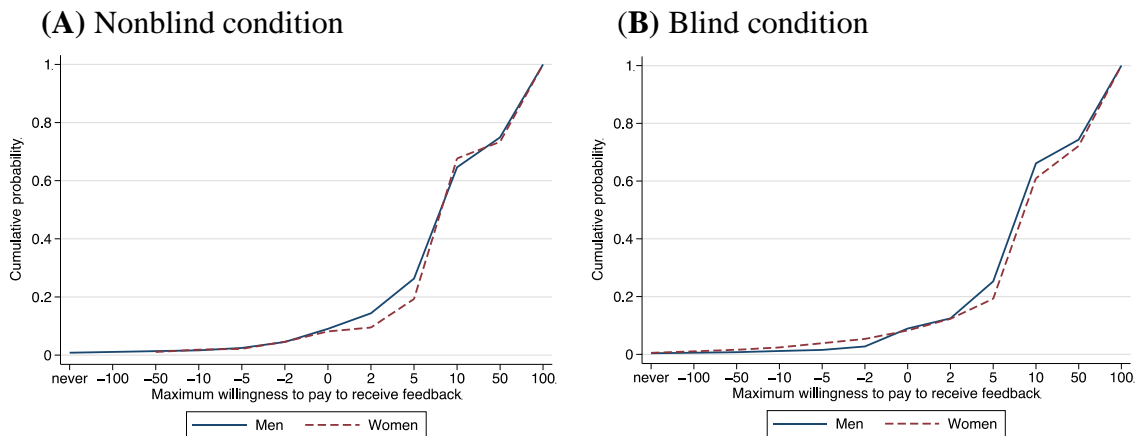


Figure 3. Maximum Willingness to Pay for Feedback on Interview Questions

Notes: Observations from the Interview Preferences for Feedback Study. Sample restricted to subjects with monotonic preferences for feedback over all prices.

¹⁸ Table B5 shows that failing monotonicity is significantly associated with being male, finding the instructions difficult, having lower AI score, and reporting lower interest in receiving feedback in the unincentivized question, and not associated with treatment assignment.

Figure 3 shows the cumulative distribution of the maximum a subject is willing to pay to receive feedback.¹⁹ The majority of participants (91.5 percent) are willing to pay a strictly positive amount. The median and modal willingness to pay is 10 sliders. Strikingly, 26.5 percent of participants chose to receive feedback irrespective of the price. Figure 3 shows little differences in the distributions across treatment or gender.

Table 3. Maximum willingness to pay to receive interview feedback when it is costly

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	3.523 (3.173)	0.713 (3.179)	-0.577 (4.478)	0.377 (3.159)	1.312 (3.146)	-1.160 (3.128)	0.456 (3.139)
Blind	2.527 (3.118)	2.859 (3.074)	1.359 (4.837)	2.973 (3.065)	2.492 (3.043)	1.940 (3.001)	2.153 (3.030)
Female x Blind			2.544 (6.314)				
Average AI z-score				4.296* (2.365)	4.571* (2.349)	4.345* (2.327)	4.713** (2.346)
Word count				0.037** (0.016)	0.043*** (0.016)	0.043*** (0.016)	0.042*** (0.016)
Belief of rank (1:best , 10:worst)					-1.481 (0.931)	-1.937** (0.919)	-1.829** (0.927)
Certainty in rank					5.791*** (1.663)	3.554** (1.683)	4.585*** (1.680)
Positive influence						10.244**** (1.697)	
Negative influence							5.954**** (1.647)
Controls	N	Y	Y	Y	N	Y	Y
N	1,309	1,309	1,309	1,309	1,309	1,309	1,309

Notes: Observations from the Interview Preferences for Feedback Study. Coefficient estimates from interval regressions of the maximum willingness to pay to receive feedback when it is costly to do so (i.e., when the price to receive feedback is 2, 5, 10, 50, or 100 sliders). *Positive influence* (*Negative influence*) is the first component derived from principal component analysis of the three questions on the perceived influence of better (worse) than expected feedback. Controls are age, race, region of residence, educational attainment, high school in the US, 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. Columns 6-7 control in addition for the order of the questions on the influence of receiving better or worse than expected feedback. Sample restricted to subjects with monotonic preferences for feedback over positive prices. Robust standard errors in parentheses. *p<0.1, **p<0.05, ***p<0.01, ****p<0.001.

¹⁹ 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.

Table 3 looks at the factors that predict the demand for feedback and asks whether conditioning on these factors changes the null gender difference. As in the Cognitive Test study, we focus 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—since variability is greatest over positive prices. The sample in the analysis in Table 3 consists of subjects with monotonic preferences for feedback over positive prices, which is 97 percent of subjects in Session 2. Table 3 shows coefficient estimates from interval regressions of the maximum willingness to pay under several specifications. The maximum willingness to pay is not significantly associated with gender in any specification. We estimate that the average gender difference is approximately 1 to 3.5 sliders (women are willing complete approximately 1 to 3.5 more sliders than men to receive feedback), a value statistically indistinguishable from zero. 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.

We also find no difference across treatment, neither overall nor for women specifically. This suggests that introducing the possibility of gender discrimination does not produce a gender gap in demand for feedback. In the Appendix, we explore how experiences and expectations of discrimination impact the demand for feedback. While women report worrying about sex-based discrimination to a greater extent than men, expectations of discrimination do not predict demand for feedback in either treatment, nor do they influence the estimated gender gap.

Turning to other factors, our results are largely consistent with the Cognitive Test Study. The number of words written in total across the three interview questions is predictive of greater willingness to pay ($p=0.017$ in column 4 and $p=0.007$ in columns 5-7), perhaps reflecting an association between effort in answering the questions and greater demand for feedback.

Participants with higher objective scores have greater willingness to pay ($p=0.045$ to $p=0.069$ in columns 4-7), which may reflect an association between performance and demand for feedback and/or an association between personality (intellectual curiosity, a tendency to strive for achievement, assertiveness, and tolerance to stress) and demand for feedback. Specifications 5-7 in Table 3 show that participants who believe they ranked better ($p=0.035$ to $p=0.112$) and who are more certain of their beliefs ($p<0.001$ to $p=0.035$) are willing to pay more for feedback. Table B7 replicates specification 5 in Table 3, but now including as additional control the interaction between beliefs and certainty of beliefs. We find a negative and significant interaction effect, indicating that individuals who are certain of good performance have in particular greater demand for feedback, suggestive of a consumption channel for demand for feedback.

Turning our attention to self-reported expected impact, we find that women expect that feedback will be more informative for their evaluation of their performance, their ability to perform well in job interviews generally, and their capabilities in other aspects of life (Table B8). This is true both about receiving feedback better and worse than expected. Treatment has no significant impact on these self-reports, and there is no gender-treatment interaction (Table B9). These results are somewhat in contrast to the Cognitive Test Study, where women expected the feedback to be somewhat more informative in terms of that specific task, but less influential in a generalized way. This could reflect task differences or differences in the nature (e.g., subjectivity) of the feedback; understanding what predicts gender differences in these types of perceptions would be a fruitful avenue for future work.

Columns 6-7 of Table 3 augment specification 4 in Table 3 by including the first principal component of each set of three variables.²⁰ These columns show that greater self-reported

²⁰ As in the Cognitive Test Study, responses to the three variables that involve better-than-expected feedback are highly correlated with each other and the first principal component derived from PCA alone explains approximately

influence is significantly associated with higher willingness to pay to receive feedback, consistent with the Cognitive Test Study. But, despite women reporting greater influence of feedback in this environment, we continue to find no significant gender differences in demand for feedback.

In sum, we do not find that women are more or less averse to feedback than men in our environment. This is also the case for demand for feedback at zero price (Table B11).

4.3.3 Beliefs of Preferences for Feedback

We now turn attention to beliefs about demand for feedback. To mirror the analysis in the previous section, we use participant beliefs about how many of 100 men (women) demand feedback at different prices to construct implied beliefs of maximum willingness to pay. We do this using only observations from participants who provided monotonic beliefs over positive prices, 86.7% of the sample. Table 4 presents the results, using interval regressions to predict the believed maximum willingness to pay.

Specification 1 in Table 4 performs a within-subject analysis, regressing the maximum willingness to pay on an indicator that the belief is about women and includes subject fixed effects. We estimate that subjects anticipate women are willing to pay 0.73 fewer sliders than men ($p < 0.01$). Specification 2 performs an across-subjects analysis, using only the first set of beliefs 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 anticipated gender gap: subjects anticipate women are willing to pay 4.16 fewer sliders than men ($p = 0.018$). Specification 3 shows that controlling for the belief provider's gender does not eliminate the believed gender gap. Table B13 in the Appendix shows that the believed gender gap in demand for feedback is

88 percent of the variation in these variables. This is also the case for the three variables that involve worse-than-expected feedback.

indistinguishable across the blind and nonblind conditions, and that both men and women believe that women have lower demand for feedback. In Table B14, we also document a believed gender gap in demand for feedback when we examine beliefs about demand for feedback at zero price over the entire sample of subjects.

Table 4. Beliefs of maximum willingness to pay to receive costly feedback on interview performance

	(1)	(2)	(3)
Female	-0.728*** (0.224)	-4.158** (1.764)	-4.151** (1.777)
Blind		2.007 (1.998)	1.949 (1.993)
Female self			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 Beliefs Study. Coefficient estimates from interval regressions of the maximum willingness to pay to receive feedback when it is costly to do so (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 subjects with monotonic beliefs about preferences for feedback over positive prices. Each subject generates a set of 100 observations corresponding to the subject's beliefs of the feedback choice of 100 males, and a set of 100 observations corresponding to the subject's beliefs of the feedback choice of 100 females. Column 1 uses the full set of 200 observations per subject. Columns 2-3 use 100 observations per subject, corresponding to the belief about the gender first elicited from the subject. Robust standard errors in parentheses in column 1, and clustered at the subject level in columns 2-7. *p<0.1, **p<0.05, ***p<0.01, ****p<0.001.

Finally, we examine our more qualitative questions. Participants indicate that they believe women are less likely to demand performance feedback than men for the interview task specifically; 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 (t-test $p=0.034$). On the other hand, for tasks more generally, participants 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 $p=0.435$).²¹ Note that this is in contrast to what we found for the *identical* generalized question from participants assigned to the cognitive skills test version of the study. It's 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 a mock job interview, whether the feedback is objective or subjective, and whether it is possible for the provider of the feedback to discriminate on the basis of the participant's gender. Across these variations, we find that women demand performance feedback no less than men.

We find that the participants' perceptions of how informative the feedback will be for self-assessment and their expectations of receiving good news both help to explain their demand for feedback. Consistent with past work, women in our two studies are less confident in their performance than men and perceive feedback to be more influential and generalizable for self-assessment than men do, results that have been hypothesized in the literature to imply that women will shy away from performance feedback more so than men. Yet, we see no evidence that women are more feedback averse than men even in spite of these gender differences in other dimensions.

Our results suggest that gender differences in demand for performance feedback may not be a large factor in driving differences in educational and labor market outcomes. Of course, we

²¹ Table B15 shows that these beliefs vary by gender. Men anticipate women to be more feedback averse on the interview task and on tasks more generally, while women anticipate no gender differences in either. Men's mean answer is 0.193 ($p=0.002$) for the interview task and 0.144 ($p=0.014$) for tasks more generally, while women's mean answer is -0.01 ($p=0.926$) for the interview task and -0.087 ($p=0.185$) for tasks more generally.

cannot rule out that women may indeed be more averse to receiving performance feedback in other settings. Future work should 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.

While we find no evidence of a gender gap in demand for feedback, we do observe that individuals anticipate women being less eager to receive performance feedback than men. These inaccurate beliefs could drive under-provision of feedback to women. More broadly, organizations should consider how to structure processes to ensure equal access to high-quality, well-delivered feedback, further increasing the chances that differences in feedback provision do not contribute to gender gaps in career advancement.

REFERENCES

- Araujo, F.A., Carbone, E., Conell-Price, L., Dunietz, M.W., Jaroszewicz, A., Landsman, R., Lame, D., Vesterlund, L., Wang, S.W., and Wilson, A.J., 2016. The slider task: An example of restricted inference on incentive effects. *Journal of the Economic Science Association*, 2: 1-12.
- Ashford, S.J., Blatt, R., and Vandewally, D., 2003. Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of Management*, 29.6: 773-799.
- Bolte, L., and Raymond, C.B., 2022. Emotional inattention. *Working paper*.
- Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A., 2019. Beliefs about gender. *American Economic Review*, 109.3: 739-773.
- Burks, S.V., Carpenter, J.P., Goette, L., and Rustichini, A., 2013. Overconfidence and social signalling. *Review of Economic Studies*, 80.3: 949-983.
- Buser, T., Niederle, M., and Oosterbeek, H., 2014. Gender, competitiveness, and career choices. *Quarterly Journal of Economics*, 129.3: 1409-1447.
- Coffman, K., 2014. Evidence on self-stereotyping and the contribution of ideas. *Quarterly Journal of Economics*, 129.4: 1625-1660.
- Coffman, K., Collis, M., and Kulkarni, L., 2020. Stereotypes and belief updating. *Working paper*.
- Coffman, K., and Klinowski, D., 2020a. A study of feedback aversion. *AEA RCT Registry*, June 11. <https://doi.org/10.1257/rct.5990-1.0>
- Coffman, K., and Klinowski, D., 2020b. A study of feedback aversion on a job interview task. *AEA RCT Registry*. November 13. <https://doi.org/10.1257/rct.6735-1.0>
- Coffman, K., and Klinowski, D., 2022. Beliefs about receiving feedback. *AEA RCT Registry*. November 18. <https://doi.org/10.1257/rct.10413-1.0>
- Danz, D., Vesterlund, L., and Wilson A.J., 2020. Belief elicitation: Limiting truth telling with information on incentives. *NBER Working Paper 27327*,
- De Quidt, J., Vesterlund, L., and Wilson, A.J., 2019. Experimenter demand effects. In *The Handbook of Research Methods and Applications in Experimental Economics*, eds. Schram, A., and Ule, A., Edward Elgar Publishing.
- Dweck, C.S., 2000. *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

- Eil, D., and Rao, J.M., 2011. The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3.2: 114-138.
- Eliasz, K., and Spiegel, R., 2006. Can anticipatory feelings explain anomalous choices of information sources? *Games and Economic Behavior*, 56: 87-104.
- Ganguly, A., and Tasoff, J., 2017. Fantasy and dread: The demand for information and the consumption utility of the future. *Management Science*, 63.12: 4037-4060.
- Gill, D., and Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102.1: 469-503.
- Golman, R., and Loewenstein, G., 2018. The demand for, and avoidance of, information. *Working paper*.
- Grosse, N.D., and Reiner, G., 2010. Explaining gender gaps in competitiveness: Gender-task stereotypes. *Working paper*.
- Healy, P.J., and Kagel J., 2022. Testing elicitation mechanisms via team chat. *Working paper*.
- Karlsson, N., Loewenstein, G., and Seppi, D., 2009. The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, 38.2: 95-115.
- Klinowski, D., 2019. Selection into self-improvement and competition pay: Gender, stereotypes, and earnings volatility. *Journal of Economic Behavior & Organization*, 158: 128-146.
- Köszegi, B., 2006. Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4.4: 673-707.
- Masatlioglu, Y., Orhun, A.Y., and Raymond, C. 2021. Intrinsic information preferences and skewness. *Working paper*.
- Mobius, M.M., Niederle, M., Niehaus, P., and Rosenblat, T.S., 2022. Managing self-confidence: Theory and experimental evidence. *Management Science*.
- Moore, D.A., and Healy, P.J., 2008. The trouble with overconfidence. *Psychological Review*, 115.2: 502-517.
- Niederle, M., and Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122.3: 1067-1101.
- Roberts, T.-A., and Nolen-Hoeksema, S., 1989. Sex differences in reactions to evaluative feedback. *Sex Roles*, 21: 11-12.
- Shastri, G.K., Shurchkov, O., and Xia L.L., 2020. Luck or skill: How women and men react to noisy feedback. *Journal of Behavioral and Experimental Economics*, 88: 101592.

Shurchkov, O., 2012. Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10.5: 1189-1213.

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

Weinberg, B.A., 2009. A model of overconfidence. *Pacific Economic Review*, 14.4: 502-515