Luck or skill: How women and men react to noisy feedback

Gauri Kartini Shastry\textsuperscript{a}, Olga Shurchkov\textsuperscript{a,}\textsuperscript{⁎}, Lingjun Lotus Xia\textsuperscript{b}

\textsuperscript{a} Department of Economics, Wellesley College, 106 Central St., Wellesley, MA, United States
\textsuperscript{b} Massachusetts Institute of Technology, Harvard University, MA, United States

\begin{abstract}
We design an experiment that sheds light on the asymmetry in how men and women interpret noisy feedback about relative standing and how this gender difference can affect tournament entry. Women attribute negative feedback to lack of ability, even when the feedback is due to bad luck. High-ability men who receive negative feedback correctly attribute it to luck. Men attribute negative feedback to lack of ability only when it confirms prior beliefs. We find consistent gender differences in tournament entry: noisy feedback eliminates the gender gap but primarily because low-performing men opt out of tournament. High-performing women who receive surprising negative feedback reduce tournament entry, generating a gender gap in performance and earnings relative to the setting without feedback.
\end{abstract}

\section{Introduction}

Gender gaps in economic outcomes are pervasive and persistent. A simple comparison of annual wages for full-time working women and men reveals a gender gap of about 21 percent (Blau and Kahn 2017). Only 5.8 percent of Fortune 500 companies have female CEOs, with women occupying a meager 26.5 percent of senior-level managerial positions (Catalyst 2020). Leadership gaps are also prevalent in politics. In the US, women hold just around 20 percent of seats in the House of Representatives and Senate. The problem is not restricted to the US: women hold only around a quarter of the seats in national parliaments worldwide (The World Bank 2019).

Explanations for these gender gaps fall into two broad categories. First, women might be discriminated against by employers, superiors, and/or coworkers, which may lead to fewer opportunities to achieve better economic outcomes even if we compare men and women within a given occupation or position (Sarsons 2017; Goldin 2014, for an overview). Second, women might self-select into lower-paying jobs and be less likely to undertake lucrative opportunities (Blau and Kahn 2017). In this paper, we focus on gender differences in behavioral traits that may serve as underlying mechanisms behind this differential sorting (see Niederle 2016 and Shurchkov and Eckel 2018 for comprehensive surveys of the literature). In particular, a seminal study by Niederle and Vesterlund (2007) – hereafter NV – shows that men are twice as likely as women to enter a tournament, even when there is no significant gender gap in baseline ability. Subsequent research has investigated the provision of feedback on relative ability as a way to “nudge” women to be more competitive (Ertac and Szentes 2011; Brandts, Groenert, and Rott 2014; Berlin and Dargnies 2016). In many real-world settings, feedback depends partly on an individual’s ability and partly on luck. This paper studies differences in how men and women attribute noisy feedback to luck and ability, and how this relates to tournament entry; we then explore how these differences can affect selection into competition and, subsequently, earnings.

In our online experiment, participants first perform a task and report beliefs about their own performance. They receive a payment based partly on their own performance (ability component) and partly on whether their performance is higher or lower than that of a given occupation or position (Sarsons 2017; Goldin 2014, for an overview). Second, women might self-select into lower-paying jobs and be less likely to undertake lucrative opportunities (Blau and Kahn 2017). In this paper, we focus on gender differences in behavioral traits that may serve as underlying mechanisms behind this differential sorting (see Niederle 2016 and Shurchkov and Eckel 2018 for comprehensive surveys of the literature). In particular, a seminal study by Niederle and Vesterlund (2007) – hereafter NV – shows that men are twice as likely as women to enter a tournament, even when there is no significant gender gap in baseline ability. Subsequent research has investigated the provision of feedback on relative ability as a way to “nudge” women to be more competitive (Ertac and Szentes 2011; Brandts, Groenert, and Rott 2014; Berlin and Dargnies 2016). In many real-world settings, feedback depends partly on an individual’s ability and partly on luck. This paper studies differences in how men and women attribute noisy feedback to luck and ability, and how this relates to tournament entry; we then explore how these differences can affect selection into competition and, subsequently, earnings.

In our online experiment, participants first perform a task and report beliefs about their own performance. They receive a payment based partly on their own performance (ability component) and partly on whether their performance is higher or lower than that of a...
randomly chosen participant or “opponent” (luck component). We assign subjects into conditions based on whether they know the gender of their random match, following the literature which points to the importance of stereotype threat in mixed gender settings for driving gender gaps in outcomes (see for example, Gneezy, Niederle, and Rustichini, 2003).

Upon seeing their payment information, subjects report their expectation on whether this payment is above or below the average payment of a comparable group of participants (we call this measure “self-evaluation”). Subjects are then randomized into two main treatment conditions: control with no additional feedback and treatment with additional feedback on how their individual payment actually compares to the average. We refer to the control group as the limited feedback condition since the payment itself may reveal noisy information about one’s ability.

The first novel feature of our design is that we ask the subjects to assess the extent to which they believe their above- or below-average payment arose due to luck or ability (bad/low if the payment is below average or good/high if the payment is above average). This means that feedback in our experiment can be negative/positive and surprising/confirming, which allows us to study the gender differences in attribution of reinforcing feedback and unexpected feedback. We find that men attribute negative feedback to low ability, regardless of their actual ability. On the other hand, men of high ability who receive negative feedback by chance correctly attribute it to luck. Men attribute negative feedback to low ability only when the feedback confirms a pre-existing negative self-evaluation.

Our next contribution is to show that this asymmetry in attribution of feedback may contribute to gender differences in real-world relevant outcomes, such as willingness to compete. After the attribution question, subjects perform the same task again, but first choose which payment scheme to apply to their second round performance (piece-rate or tournament). Without additional feedback, we document a significant gender gap in tournament entry, particularly in the treatments where the gender of the random match is unknown or the opponent is male (consistent with NV and the subsequent large literature on competitiveness, see Niederle 2016 for a survey).

Additional feedback eliminates this gap, which is consistent with previous findings (e.g., Brandts, Groenert, and Rott 2014; Wozniak, Harbaugh, and Mayr 2014). In our experiment, however, feedback erases the gender gap in tournament entry by correcting the suboptimal entry pattern of low-performing men who now choose not to enter competition, a finding consistent with previous studies that find that men are more likely to revise their beliefs about performance as compared to women who receive similar feedback (for example, Mobius et al. 2011; Brandts, Groenert, and Rott 2014). Specifically, men positively sort into tournament only with additional feedback. Women, on the other hand, select into tournament based on performance even with limited feedback. The gender gap in tournament entry we document in our limited feedback treatment is not driven by high-ability women shying away from competition, unlike previous literature which finds that the gender gap in competitiveness increases with ability (NV; Buser, Peter and Wolter 2017; Buser, Ranellih and van Veldhuizen 2017). On average, additional feedback eliminates the gap by altering the behavior of men.

This average effect, however, masks the heterogeneity in the impact of positive and negative feedback. While we find that women sort positively into tournament in both treatment conditions, on average, when we break up our data by positive and negative feedback, we find that negative feedback eliminates the positive selection of women into tournament. Specifically, high-ability women who receive negative feedback opt out of the tournament relative to comparable women in the limited feedback condition.

We provide a novel explanation for the patterns in tournament entry in response to feedback: gender differences in attribution. Low-ability men who would have entered the tournament without feedback, receive a negative signal and attribute it to low ability in the task. At the same time, high-ability men who receive a negative signal correctly attribute it to bad luck and continue to enter the tournament. Women, on the other hand, attribute negative feedback to lack of ability, regardless of whether it is consistent with their self-evaluation or actual ability. Thus, high-ability women opt out of competition after receiving negative feedback, even though negative feedback for them is primarily due to bad luck. Women are also more likely than men to attribute surprising positive feedback to luck, which could explain why we do not find evidence of high-ability women increasing tournament entry after receiving positive feedback although we have limited statistical power when studying the impact of unexpected positive feedback.

Our findings imply that the asymmetric way in which men and women attribute feedback on relative standing may have welfare consequences. On one hand, because of how men attribute noisy feedback, feedback can increase efficiency by reducing inefficient entry of low-performing men into the tournament. On the other hand, noisy feedback can also decrease efficiency by reducing entry of high-performing women who get noisy feedback by chance and blame themselves for the bad outcome. In fact, we find that feedback decreases the performance of high-ability women whether or not they enter the tournament, without much of an effect on low-ability women. We also find that the noisy feedback generates a gender gap in earnings, because it leads high-ability women to forego potentially higher payoffs that they could have earned if they were to compete. We do not take a stand on which of these effects would dominate, but simply point out that noisy feedback may exacerbate gender differences in outcomes.

This paper relates to several literatures. One contribution of this paper is to show that gender differences in attribution of noisy feedback to luck and ability can produce gendered effects on tournament entry, particularly for high-ability women. The question of how people attribute successes and failures to ability or luck has been studied extensively in social psychology revealing that people exhibit “self-attribution bias:” the tendency to attribute success to own ability, but failures to external factors (lack of ability), while men are more likely to attribute “bad news” feedback to external factors (bad luck). Our findings of gender differences in attribution of feedback provide a novel explanation for recent findings of gender differences in tournament entry even after receiving positive feedback. For example, Buser and Yuan (2019) show that women are less likely than men to compete after losing in an earlier round. Along the same lines, Filippin and Gioia (2018) find that men become significantly more risk averse after losing a tournament than after randomly earning the same low payoff. Our findings are consistent with this observation, as men are more likely to blame the loss on bad luck instead of ability, increasing their risk-aversion.

Our second contribution is to use our result on the asymmetry of attribution by gender to shed light on the persistent gender differences in competitiveness. Importantly, these differences are driven by the gender stereotypes associated with a given environment: women are less competitive only in male-typed settings (Günther et al. 2010; Shurckov 2012). We focus our attention specifically on such a context, the mental rotation task (MRT), in order to model the relevant gender gap in the choice to study stereotypically male subjects in college (STEM) or pursue careers in male-dominated industries. For example,
building on NV, Buser, Niederle, and Oosterbeek (2014) show that the gender difference in tournament entry explains a substantial portion of the gender difference in academic track choice.

Our study also contributes to a growing literature dedicated to investigating how individuals respond to feedback in terms of changing their beliefs or choices. The former focuses on eliciting posterior belief distributions and comparing them to the Bayesian benchmark (see for example, El and Rao 2011; Buser, Gehards, and van der Weele 2018). Coffman et al. (2019) is the latest in this line of research, exploring the role of stereotypes in the way in which men and women react to noisy absolute feedback about performance. Our work shifts focus from beliefs to outcomes, by exploring the mechanisms behind the gendered effect of noisy feedback on the choice to enter a tournament. Previous experiments have shown that sufficiently precise information about one’s own and/or one’s opponents’ performances can significantly reduce or eliminate the gender gap in tournament entry (Ertac and Szentes 2011; Ewers 2012; Wozniak et al. 2014; Berlin and Dargnies 2016; and Baier et al. 2018), though this is not the case for aggregated information (Jeworrek, 2016; Wozniak et al., 2014). Even though it is possible to interpret feedback provided to subjects in some previous studies as inherently noisy (for instance, because past outcomes may involve some luck such as in Brands, Groenert, and Rott 2014 or in Buser and Yuan 2019), we bring in the luck component more explicitly and emphasize its existence to the subjects in our experiment.

The rest of the paper is organized as follows. Section 2 presents our methodology and states the ex-ante hypotheses. Section 3 describes the experimental design and discusses its limitations. Section 4 provides a first look at the data. Section 5 presents our main findings on the gender differences in attribution of noisy feedback and applies the results on attribution to the gender gap in tournament entry. Section 6 discusses the implications of the observed gender asymmetry in attribution in terms of efficiency and welfare in our setting. Section 7 concludes.

2. Methodology and hypotheses

Before we describe our procedures and treatments in greater detail, we present the hypotheses that our controlled online experiment was designed to test.

**Hypothesis 1.** Men and women attribute additional noisy feedback differently, depending on the nature of feedback:

- Overall, women are more likely to attribute negative/positive feedback to lack of ability/good luck, while their male counterparts are more likely to attribute negative/positive feedback to bad luck/high ability.
- The gender asymmetry in attribution of feedback is more pronounced when feedback (negative or positive) comes as a surprise, compared to one’s own expectation about relative performance.

Hypothesis 1 is based on the previous literature on the existence of the overall attribution bias (Hoffmann and Post 2014) and on previous accounts that show that women are likely to blame themselves for factors that are completely outside of their control, such as infertility (McLeod and Ponesse 2008).

The next two hypotheses begin to unpack how noisy feedback can impact the gender gap in tournament entry. To start, Hypothesis 2 tests the standard finding of a gender gap in competitiveness in the literature, establishing a baseline condition (Niederle and Vesterlund 2007).

**Hypothesis 2.** With limited feedback about relative performance, women are less likely to enter a subsequent tournament than their male counterparts, particularly when the gender of the opponent is either unknown or known to be male than when the opponent is known to be female.

The second part of Hypothesis 2 is based upon the previous observation that gender differences in competitiveness may be attenuated in same-sex competitions (see for example Niederle and Vesterlund 2011).

While there is no clear agreement in the literature about the effect of information on tournament entry, we rely on the majority of studies that show that feedback can mediate the gender gap in competitiveness.

**Hypothesis 3.** The gender gap in tournament entry is significantly reduced in the additional feedback condition relative to the baseline treatment with limited feedback.

Finally, Hypothesis 4 elaborates on the effects of attribution bias on the gender gap in the tournament entry decision.

**Hypothesis 4.** Gender asymmetry in attribution produces differential effects of additional feedback about relative standing on tournament entry:

- Women who receive surprising negative feedback are more likely to shy away from competition as compared to men in the same group and as compared to women who did not receive such feedback.
- Men who receive surprising positive feedback are more likely to enter competition as compared to women in the same group and as compared to men who did not receive such feedback.
- Subjects who receive feedback confirming their prior expectation (negative or positive) do not differ in their tournament entry decision from subjects who did not receive such feedback.

In order to test the above hypotheses, we design a controlled online survey experiment that establishes an environment where we can observe the effect of noisy feedback on entry into competition and on how men and women differ in their attribution of that feedback. Our experiment also generated a rich set of data that allows us to explore a variety of additional channels and patterns that we may not have expected ex ante. For example, it is likely that negative surprises would occur more often among high-performers or those with high absolute confidence levels, while positive surprises would likely occur among low-performers or those with low absolute confidence levels. Therefore, we will explore the relationship between ability and tournament entry for men and women.

3. The online experiment

3.1. Baseline ability and confidence in the Mental Rotation Task (MRT)

In round 1 of the experiment, subjects began by solving Mental Rotation Task (MRT) questions. In an MRT question, participants see a target three-dimensional shape made of 10 cubes and are asked to identify the rotated version of the target shape among the three choices. Fig. 1 shows a sample problem.

We used pilot surveys to confirm that the MRT requires skill unlike the “find the median task” and is perceived to favor men unlike the “pattern” task, as we intended. Furthermore, out of the three suitable tasks we considered, the MRT has been previously used to study gender differences in performance in competition (Iribarri and Rey-Biel 2017). Online Appendix C provides a more detailed description of the task and the considerations that led to its selection.

Subjects solved as many MRT questions as possible out of 8 in 2.5 minutes, receiving one point for each correct answer and zero points for each wrong or blank answer (piece-rate payment scheme). The total number of points represents one’s score. After completing the task, participants estimated their score, and the resulting value measures each subject’s score confidence.

---

Our task being male-favoring could explain why women attribute negative feedback to ability. Coffman et al. (2019) find that men attribute negative feedback even more to ability in female-favoring tasks. We chose a male-oriented task, because we were interested in gender gaps in competitiveness which arise most often in male-dominated domains, and because we were motivated by gender gaps in educational and occupational choices.
3.2. Payment information and self-evaluation

Our subjects were not informed about their actual score at any point in the experiment. Instead, subjects learned their resulting payment (in $). They were informed that, in order to compute their payment, the computer randomly selected another participant who had previously taken the same test and compared their two scores. If the participant’s score was higher than or equal to that of her random match, she got 20 cents for each correct answer. However, if her score was lower than that of her random match, she got only 15 cents for each correct answer. (We discuss the effects of divisibility differences across these two outcomes in Section 3.4.)

Upon receiving payment information, all subjects were asked to assess whether they believed that their payment fell above or below the average payment earned by a previous group of participants (wave 2 of the pilot, see online Appendix C for details). This measure allows us to gauge whether a given subject has a favorable or a negative view of their outcome. Specifically, we say that a subject who perceives her payment as below average has a “negative self-evaluation,” while a subject who perceives her payment as above average has a “positive self-evaluation.”

3.3. Randomization into treatments and attribution

Our primary treatment variation is whether or not participants receive additional information about their relative performance on the MRT task.

In the Limited feedback condition (LF) participants did not receive any additional information. We refer to this treatment as the limited feedback condition, because the payment itself can be construed as a noisy signal about ability. What is important is that participants did not receive any further information with which to benchmark the payment.

In the Additional feedback condition (AF), participants learned whether their payment was actually above or below average. Note that, while the payment was calculated based on performance relative to a specific randomly matched individual, this additional feedback is benchmarked to the average of all previous subjects, importantly, the pool of possible random matches for subsequent tournament rounds.

Next, subjects were asked to attribute the relative importance of luck, as opposed to their own ability, in determining their payment outcome (see Fig. 2). The resulting attribution measure allows us to directly test Hypothesis 1.

Note that the nature of the attribution question differs across our two feedback treatments. Participants who did not receive the information on whether their payment was above/below average (LF treatment) assessed the contribution of luck and skill to their payment purely based on their perception of relative payment – their self-evaluation. A participant with a positive/negative self-evaluation assessed to what extent her presumed above-average/below-average payment was due to her own high/low ability, and to what extent it was due to good/bad luck of being matched with a relatively weak/strong participant from the other group. On the other hand, participants who learned whether their payment was actually above or below average (AF treatment) assessed the contribution of luck and skill to their payment based on their true standing relative to the average.

In addition to the feedback treatments, we varied the information about the gender of the random match whose performance was used for comparison in the payment calculation. This treatment dimension explores whether the salience of opponent gender affects subsequent competitive choices of male and female subjects, building on previous literature that suggests that group composition can affect performance under competition (Gneezy, Niederle, and Rustichini 2003) and tournament entry decisions (Datta Gupta, Poulsen, and Villeval 2013).

In the Match with unknown gender (UG) condition, we used a gender noncommittal pronoun “his/hers” in the explanation of how payment was calculated. In the Match with known gender (KG) condition, the payment information text subtly revealed the actual gender of the randomly selected match. We kept this reference subtle in order to avoid alerting the participant to the fact that the research questions related to gender.

Table 1 reviews the $2 \times 2$ factorial design in our experiment and

---

3 The random match was drawn from the pool of participants in our pilot survey, wave 2. Subjects in this wave of the pilot were paid in an analogous manner to those in our main experiment. See online Appendix C for detailed information on pilot waves 1 and 2.
We also ran a third treatment condition where participants (101 in total) wished to apply to subsequent performance. The two choices were:

- Set of MRT problems in 2.5 minutes, subjects chose the payment scheme to round 2 of the experiment – the choice round. Before solving another
- Tournament entry

provides the total number of subjects in each treatment condition.\(^4\)

### 3.4. Tournament entry

After the subjects answered the attribution question, they moved on to round 2 of the experiment – the choice round. Before solving another set of MRT problems in 2.5 minutes, subjects chose the payment scheme they wished to apply to subsequent performance. The two choices were:

- **Piece rate**: Participant gets $0.175 for each correct answer and 0 otherwise, regardless of anyone else’s score.
- **Tournament**: We randomly match the participant with another participant from pilot wave 2. Scoring higher than the random match yields $0.25 for each correct answer, while scoring lower yields $0.10 for each correct answer.\(^5\)

The rates at which men and women choose the tournament payment scheme, overall and when facing known- and unknown-gender opponents, allow us to test Hypothesis 2. Comparing the gender differences in tournament entry with and without additional feedback addresses our Hypothesis 3. Finally, decomposing tournament entry of men and women by different types of feedback (negative/positive) provides us with a test of Hypothesis 4.

\(^4\) We also ran a third treatment condition where participants (101 in total) indicated their willingness to pay to receive feedback on relative payment (see Appendix B for details). Unlike Wozniak et al. (2016), where the demand for relative feedback was quite high, very few participants were interested in receiving feedback on relative payment. This may be due to our feedback being intentionally uninformative. Because the option of paying for feedback makes it difficult to compare those who sorted into receiving feedback to those who did not, we omit the optional feedback condition data from all subsequent analysis.

\(^5\) Note that our experiment departs from the winner-take-all tournament design adopted in most previous studies on competitiveness, including NV. We believe that our design better approximates the nature of most competitions, where the loser still walks away with a prize, albeit a more modest one. For example, the candidate who applies for but does not get the promotion is unlikely to be fired and can still go back to the original job.

### 3.5. Procedures and limitations

The experiment was programmed in Qualtrics and conducted using the Amazon Mechanical Turk (AMT) platform between November 2017 and September 2018.\(^6\) The experiment started with informed consent. Consenting subjects read onscreen instructions and had a chance to practice with the task. We included checks throughout the experiment to ensure attention. At the end of the experiment, each participant filled out a short questionnaire which included demographic questions, a self-reported risk preference measure, and a question to elicit gender perception associated with the MRT (consent form, instructions, and the questionnaire can be found in Appendix A). Table F1.1 in the online appendix reports summary statistics of demographics and performs balance tests between treatment groups. Only 2 out of 71 comparisons across treatments are significant at the 0.05 level. All the procedures are randomized in the survey, but we confirm that results are robust to the inclusion of demographic controls (see online Appendix G). Men and women in our sample are statistically indistinguishable on most demographic characteristics (see online Appendix Table F1.2). Including the base payment of $0.50, the average payment was $2.1. The maximum payment was $4.1. The average duration of the experiment was about 11 minutes.

We conducted four waves of data collection in the main experiment over the course of November 2017 – September 2018. In the first wave, we implemented the full experimental design as described above. AMT recorded 308 valid responses with only 1 participant dropping due to failing to pass our attention check. In the second wave, we collected 88 more observations in the AF treatment. The third wave excluded the attribution question, in light of the potential concern that it may prime tournament entry, collecting a total of 198 valid responses across the LF and the AF treatments. The fourth wave collected more data in the LF and AF treatments and, like the first and second waves, included the attribution question.

There are certain limitations in our design that led to a non-standard data collection process. First, after the first wave, we discovered a coding error that resulted in some participants receiving negative feedback when they should have received positive feedback, based on their payment relative to the average. The error was corrected for all subsequent waves. Ex post, we discovered that most of the negative

\(^6\) Workers on AMT have been shown to exhibit similar behavioral patterns and pay attention to the instructions to the same extent as traditional subjects (Paolacci, Chandler, and Ipeirotis, 2010; Germine et al. 2012). Rand (2012) reviews replication studies that indicate that AMT data are reliable. There is a growing number of studies that use AMT specifically for studying gender differences in competitiveness. For example, Apicella, Demiral, and Møllerstrom (2017) show that lab results on gender gaps in tournament entry replicate with workers on this online platform, albeit with a different task.
surprises were generated due to the randomness of being in the first wave with the coding error. Note that this was not by design; we did not set out to deceive our subjects.

Second, due to the majority of the negative surprises coming from the first wave and the small sample size from the third wave, we are unfortunately unable to test whether asking the attribution question affects many of our results; we leave the question of whether posing the attribution question primes tournament entry to future work. We note, however, that tournament entry patterns did not differ significantly in wave 3; appendix F2 goes into more detail. Furthermore, no analysis was performed during the time that data were collected from waves 1-3, which is why we proceed to pool the data from all three waves.

Finally, we added a fourth wave with the hope of collecting observations with more ‘positive surprises’ – low-performing individuals getting positive feedback – but realized ex-post that there was insufficient noise in our design to generate many surprises without the coding error, which of course could not be replicated. With these limitations in mind, we choose to report the results using all our data, and to include wave fixed effects in the analysis in order to eliminate any bias that might arise due to the timing of the waves. Note that all of our results hold when we omit wave 4 from the analysis (see Appendix D), and when we only use waves 1 and 4 (see Appendix E).

Another design limitation is that savvy participants could have realized that they had ‘won’ the first-round match-up if their payment was only divisible by 0.20 and not 0.15 (and vice-versa), as well as determining their actual score. First, note that while this would eliminate some of the uncertainty about their own ability, it would not perfectly predict their score in a future round, or whether they would win or lose in a second-round tournament. Our treatment variation is in terms of feedback about relative payment compared to the average, which means that winning/losing against one random opponent in round 1 does not guarantee being above/below average. Second, we find no evidence that men or women responded to whether or not their payment was divisible by 0.15 either in terms of differences in self-evaluation or tournament entry (see online Appendix F3 for a detailed discussion).

4. A first look at the data

Table 2 reports summary statistics of key experimental variables by gender. All of these variables are from the round 1 task, collected before subjects were randomized into feedback treatment conditions, except for self-reported risk preference, which is collected after the round 2 task. We pool the data across the four waves and across treatments. The order of variables reported in the table follows the chronological order of the design of the experiment.

We find that men outperform women in the MRT task and consequently earn a higher bonus in round 1. Indeed, there is a consensus in psychology that MRT consistently elicits gender differences in performance (Masters and Sanders 1993). See online Appendix Figure F1.1 for the distribution of performance by gender.

Women in our experiment are significantly less confident than men, both in terms of their prediction of round 1 score (score confidence) and in terms of whether they believe their payment to be above or below average (self-evaluation about relative payment).\(^7\)

\[\text{footnote (continued)}\]

We acknowledge that even payments divisible by both 0.15 and 0.2 would not be ambiguous for a participant who knew exactly how many questions they got right. Since only 20% of participants were able to correctly guess how many questions they got right and 92% of those participants had answered more questions, we think that it is unlikely that participants deduced their score.

\* Online Appendix F4 further explores the gender differences in relative confidence and self-evaluation. First, we note that the gender differences in score confidence and self-evaluation in Table 2 are robust to the inclusion of wave fixed effects and controls for the gender of opponent treatment as well as score in round 1 (Table F3.1). Second, we compare actual and predicted scores 3, which is why we proceed to pool the data from all three waves.

\[\text{footnote continued)}\]

Table 2 documents the well-established gender gap in risk preferences found in the literature, with women reporting significantly lower risk attitudes (see Shurchkov and Eckel 2018 for a comprehensive review; but see Crosetto and Filippin 2017 for an important caveat that gender gaps in risk-taking vanish in the absence of a clear safe option).

5. Results

5.1. The gender gap in attribution of noisy feedback

We begin the discussion of our main results by testing Hypothesis 1 which conjectures that men and women attribute additional noisy feedback differently, depending on the nature of the feedback. Recall that, in order to correctly measure attribution, we define four types of feedback outcomes: (1) positive feedback following a positive self-evaluation (i.e., positive reinforcement); (2) positive feedback following a negative self-evaluation (i.e., positive surprise); (3) negative feedback following a negative self-evaluation (i.e., negative reinforcement); (4) negative feedback following a positive self-evaluation (i.e., negative surprise). Furthermore, we restrict our attention to the additional feedback condition, where subjects attribute actual information on relative payment to luck and to own ability, and to waves 1, 2 and 4, as the attribution question was omitted from wave 3. The attribution measure ranges from 0 (attributing the feedback on relative payment completely to luck) to 100 (attributing the feedback completely to own ability).

Across all types of feedback, men and women do not significantly differ in their attribution, with men attributing 66.9 percent of the outcome to their own ability and women attributing 68.5 percent of the outcome to their own ability, on average. Table 3, however, decomposes average attribution by gender as well as the type of feedback subjects receive, finding partial support for Hypothesis 1. In particular, the largest gender gap in attribution exists for negative surprises (top row). Men who originally hold a positive self-evaluation (i.e., originally believed they had an above-average payment) but receive negative feedback are significantly more likely than women to attribute it to bad luck. The attribution gap is reversed when feedback reinforces the self-evaluation as positive: now men are significantly more likely than women to take credit for it as being due to their high ability (second row of Table 3).

However, Hypothesis 1 is not confirmed in the cases when subjects hold a negative self-evaluation (i.e., originally believed they had a below-average payment). Contrary to our expectation, men and women

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score in Round 1</td>
<td>4.32</td>
<td>3.77</td>
<td>0.549***</td>
</tr>
<tr>
<td>Average bonus in Round 1</td>
<td>0.83</td>
<td>0.71</td>
<td>0.122***</td>
</tr>
<tr>
<td>Score Confidence</td>
<td>4.32</td>
<td>3.61</td>
<td>0.716***</td>
</tr>
<tr>
<td>Proportion self-evaluating below average</td>
<td>0.46</td>
<td>0.63</td>
<td>-0.173***</td>
</tr>
<tr>
<td>Self-Reported Risk Preference</td>
<td>5.68</td>
<td>4.57</td>
<td>1.111***</td>
</tr>
</tbody>
</table>

Number of obs. 337 352

Notes: Significance levels *** p < 0.01, ** p < 0.05, * p < 0.1. Summary statistics are based on data from no feedback and forced feedback conditions. Risk elicitation occurred as part of the post-experiment questionnaire and therefore took place post-treatment.

Finally, Table 2 documents the well-established gender gap in risk preferences found in the literature, with women reporting significantly lower risk attitudes (see Shurchkov and Eckel 2018 for a comprehensive review; but see Crosetto and Filippin 2017 for an important caveat that gender gaps in risk-taking vanish in the absence of a clear safe option).
Results in Table 4 are qualitatively robust to including demographic controls subsequent analysis, while still including gender knowledge treatment choose to pool the data across the gender of match treatments in all reduction in the gender gap by increasing tournament entry by women (e.g., different than in the literature. While other studies find that feedback diction in Hypothesis 3. However, the driving force behind the effect isponent is known to be male or unknown. In this condition, only 28 percent of women compete, compared to 49 percent of men. These results mask heterogeneity in women’s response to the type of feedback. This test marginally rejects the null of no gender gap in Column 2 (p-value of 0.112) but not in Column 5 (p-value of 0.646). But we also test the equality of the gender gap across the limited feedback and additional feedback conditions and find p-values of 0.116 (comparing Columns 2 and 5) and 0.12 (comparing Columns 3 and 6).

We also investigate whether the tournament entry decision under the two feedback conditions interacts with ability in a nonlinear manner. Fig. 4 demonstrates this to be the case. In the absence of feedback on relative payment (Panel A), women are generally more likely to enter the tournament as their score in round 1 increases. However, for men, there appears to be no relationship between score and tournament entry; if anything, the relationship is slightly U-shaped. When feedback on relative payment is provided (Panel B), men react to it, while women appear not to react much. In particular, feedback on relative payment corrects the suboptimal entry pattern of low-performing men, but does not substantially alter the behavior of women.

While additional feedback seems to increase sorting efficiency, the results mask heterogeneity in women’s response to the type of feedback (positive or negative). This heterogeneity is important to explore, as we previously saw that attribution of feedback by men and women depends on the type of feedback they receive: women attribute surprising negative feedback to low ability, while men attribute it to back luck. Fig. 5 breaks up Panel B of Fig. 4 by the type of feedback women and men receive in the AF treatment, along the entire ability distribution. Women who receive positive feedback exhibit positive sorting based on score, but women who receive negative feedback no longer increase tournament entry as score increases. Panel B displays the relationships for men, whose tournament entry probability increases with score regardless of whether the feedback they receive is negative or positive.

Table 4 verifies that the pattern of tournament selection by gender and feedback treatment holds when we apply a linear probability model with wave fixed effects and controls for gender of the opponent condition. On average, additional feedback eliminates the gender gap in tournament entry (Column 4) relative to the limited feedback treatment (Column 1), although the p-value of this difference is 0.116. The rest of the table explores the heterogeneous effects of feedback. In particular, we ask how the effect of feedback varies with ability. We observe that, in the limited feedback condition, women with the lowest scores are marginally less likely to enter the tournament than similar men (Column 2, row 1). Scoring one point higher in round 1 is associated with an approximately five percentage point increase in the probability of entering tournament (p < 0.05) for women. However, for men, scoring one point higher is not associated with a substantial change in the probability of entering the tournament (although the interaction terms for men and women are not statistically different from each other). Column 3 shows that controls for risk preferences and confidence do not fully explain the pattern of tournament entry. These results are consistent with the finding of women acting strategically in other contexts (Exley, Niederle, and Vesterlund 2016), but is novel in the literature on competitiveness.

Columns 5 and 6 of Table 4 confirm that feedback about relative payment changes the behavior of men who now positively sort into competition in a manner similar to women. The coefficients on the male interaction terms differ significantly between Columns 2 and 5 (p-value of 0.08) and between Columns 3 and 6 (p-value of 0.03). We test for the significance of these effects in a few ways. First we conduct a joint test of the coefficient on female and the difference in the coefficients on male and female interacted with score. This test marginally rejects the null of no gender gap in Column 2 (p-value of 0.112) but not in Column 5 (p-value of 0.646). But we also test the equality of the gender gap across the limited feedback and additional feedback conditions and find p-values of 0.116 (comparing Columns 2 and 5) and 0.12 (comparing Columns 3 and 6).

We also investigate whether the tournament entry decision under the two feedback conditions interacts with ability in a nonlinear manner. Fig. 4 demonstrates this to be the case. In the absence of feedback on relative payment (Panel A), women are generally more likely to enter the tournament as their score in round 1 increases. However, for men, there appears to be no relationship between score and tournament entry; if anything, the relationship is slightly U-shaped. When feedback on relative payment is provided (Panel B), men react to it, while women appear not to react much. In particular, feedback on relative payment corrects the suboptimal entry pattern of low-performing men, but does not substantially alter the behavior of women.

While additional feedback seems to increase sorting efficiency, the results mask heterogeneity in women’s response to the type of feedback (positive or negative). This heterogeneity is important to explore, as we previously saw that attribution of feedback by men and women depends on the type of feedback they receive: women attribute surprising negative feedback to low ability, while men attribute it to back luck. Fig. 5 breaks up Panel B of Fig. 4 by the type of feedback women and men receive in the AF treatment, along the entire ability distribution. Women who receive positive feedback exhibit positive sorting based on score, but women who receive negative feedback no longer increase tournament entry as score increases. Panel B displays the relationships for men, whose tournament entry probability increases with score regardless of whether the feedback they receive is negative or positive.
The pattern is consistent with gender asymmetry of attribution: women are more likely than men to attribute negative surprises to lack of ability, which particularly affects the tournament entry decision of high-ability women who receive negative feedback by chance.

We confirm the robustness of the pattern depicted in Fig. 5 using a regression framework, where we compare men and women receiving negative feedback in the AF treatment to those who would have received negative feedback but do not by virtue of being randomized into the LF condition (see appendix F6, for estimates and discussion).

Table 4
Determinants of tournament entry decision by treatment.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Limited Feedback</th>
<th>Additional Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.115*</td>
<td>-0.244*</td>
</tr>
<tr>
<td></td>
<td>(0.0588)</td>
<td></td>
</tr>
<tr>
<td>Male x Score in Round 1</td>
<td>0.0112</td>
<td>-0.00376</td>
</tr>
<tr>
<td></td>
<td>(0.0228)</td>
<td>(0.0226)</td>
</tr>
<tr>
<td>Female x Score in Round 1</td>
<td>0.0471**</td>
<td>0.0381*</td>
</tr>
<tr>
<td></td>
<td>(0.0230)</td>
<td>(0.0228)</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>0.391</td>
<td>0.391</td>
</tr>
<tr>
<td>Observations</td>
<td>284</td>
<td>284</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0211</td>
<td>0.0358</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses. All specifications include wave fixed effects and controls for gender of the match. Columns 3 and 6 control for risk and confidence. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

The pattern is consistent with gender asymmetry of attribution: women are more likely than men to attribute negative surprises to lack of ability, which particularly affects the tournament entry decision of high-ability women who receive negative feedback by chance.

We confirm the robustness of the pattern depicted in Fig. 5 using a regression framework, where we compare men and women receiving negative feedback in the AF treatment to those who would have received negative feedback but do not by virtue of being randomized into the LF condition (see appendix F6, for estimates and discussion).

While the gender differences in tournament entry by type of feedback in Fig. 5 are consistent with the gender asymmetry in attribution, in order to test our Hypothesis 4, we have to further decompose the analysis by whether the payment feedback in the AF condition was a surprise, according to the subject’s original self-evaluation. Table 5 formally shows that differences in attribution can help us explain the gender differences in the effect of additional feedback on the tournament selection decision. Panel A uses tournament entry as the dependent variable while Panel B uses attribution. Recall that a higher value for attribution indicates more of the positive or negative feedback being attributed to own ability. When we consider the entire sample in Columns 1 and 2, we see that men and women are on average less likely to compete having received negative feedback. For those participants who initially held a negative self-evaluation (Columns 3 and 4), the effect is symmetric for both genders – receiving negative feedback decreases the likelihood of entry. Attribution can explain this: both men and women attribute reinforcing negative feedback to ability (Columns 3 and 4 in Panel B). However, the effect of negative feedback is not symmetric by gender when we consider the subsample of participants who initially held a positive self-evaluation (Columns 5 and 6). In particular, women who were surprised to receive negative feedback are significantly less likely to enter the tournament (Panel A), which could be explained by the fact that they were more likely to attribute this news to lack of ability (Panel B). On the other hand, men do not change their
tournament entry decision as much, consistent with the fact that they attribute the negative surprise to bad luck. Taken together, these results support Hypothesis 4.13

Because our previous results suggest that negative feedback interacts with ability in a nonlinear fashion (Fig. 5), we explore heterogeneity in the attribution mechanism for tournament entry by ability. In Table 6, we compare the subjects who receive negative feedback in the additional feedback treatment group and those who would have received negative feedback in the limited feedback group, breaking up the responses by ability bin to allow for the nonlinearity we saw earlier. The lowest bin includes those scoring 0 or 1 out of 8 as the omitted category, the second bin includes those scoring below average, either 2 or 3 out of 8, the third bin includes those who scored approximately average, either 4 or 5 out of 8, and the highest bin includes those score 6 or above out of 8. Even-numbered columns control for risk preference and confidence.

Columns 1 and 2 simply replicate the result that, even among those women who would have received negative feedback, there is still a positive sorting relationship with respect to score: women of higher ability are more likely to compete. However, women in the additional feedback condition (Columns 3 and 4) no longer correctly sort into tournament based on their score upon receiving negative feedback. The p-values at the bottom of the table confirm that the interaction terms are jointly significant in the limited feedback condition but not in the additional feedback condition. The sorting behavior of men does not differ substantially by feedback condition and the male interaction terms are jointly insignificant in both groups, although the interaction terms in the additional feedback condition are positive, relative to the lowest bin, and monotonically increasing (comparing Columns 3 and 4 to Columns 1 and 2).

Comparing men to women in Columns 1 - 4, we see that negative feedback discourages women, but not men, in the higher bins from entering the tournament. The p-value at the bottom of the table rejects equality of the interactions (men vs. women) in the limited feedback condition.

Columns 5 and 6 of Table 6 shed light on whether attribution can explain the gender differences. Men in the higher score bins attribute the negative feedback more to bad luck, which explains why they do

---

13 Online Appendix Tables G8 and G9 confirm that the results in Table 5 are robust to including demographic controls and using a logit specification for Panel A, respectively.

14 In the additional feedback group treatment group, 131 men and 107 women scored 4 or above, out of whom, 28 and 29 received negative feedback, respectively. See online appendix F, Table F6.2 for linear specifications that replicate the analysis in Table 6.
not alter their tournament entry behavior (the p-value at the bottom of the table indicates that the male interaction terms are jointly significant). On the other hand, whether women attribute negative feedback to ability or luck does not differ by their actual ability – in fact, women in the higher bins attribute the negative feedback more to ability than those in the bottom bin, although the differences are not

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Gender differences in the effect of receiving negative feedback on attribution and tournament entry.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Panel A: DepVar: Tournament Entry</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.0925</td>
</tr>
<tr>
<td></td>
<td>(0.0749)</td>
</tr>
<tr>
<td>Male x Negative Feedback</td>
<td>-0.195***</td>
</tr>
<tr>
<td></td>
<td>(0.0692)</td>
</tr>
<tr>
<td>Female x Negative Feedback</td>
<td>-0.311***</td>
</tr>
<tr>
<td></td>
<td>(0.0722)</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>0.381</td>
</tr>
<tr>
<td>F-test of equality of interactions (p-value)</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Observations</td>
<td>407</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0730</td>
</tr>
<tr>
<td><strong>Panel B: DepVar: Attribution</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-9.107**</td>
</tr>
<tr>
<td>Female x Negative Feedback</td>
<td>11.74***</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>67.68</td>
</tr>
<tr>
<td>F-test of equality of interactions (p-value)</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>F-test of female = female X negfbk (p)</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Observations</td>
<td>309</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0391</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses. All specifications include wave fixed effects and controls for gender of match treatment. The sample is restricted to participants in the additional feedback treatment. Panel B omits wave 3 since it did not contain the attribution question. Even numbered columns control for score, risk and confidence. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Gender differences in sorting into tournament entry in response to negative feedback.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DepVar:</td>
<td>Tournament Entry</td>
</tr>
<tr>
<td>Samples</td>
<td>Limited Feedback</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.346*</td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
</tr>
<tr>
<td>Male x Score Bin 2</td>
<td>-0.0460</td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
</tr>
<tr>
<td>Male x Score Bin 3</td>
<td>-0.110</td>
</tr>
<tr>
<td></td>
<td>(0.214)</td>
</tr>
<tr>
<td>Male x Score Bin 4</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td>(0.267)</td>
</tr>
<tr>
<td>Female x Score Bin 2</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
</tr>
<tr>
<td>Female x Score Bin 3</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
</tr>
<tr>
<td>Female x Score Bin 4</td>
<td>0.920***</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>0.341</td>
</tr>
<tr>
<td>F-test of male interactions (p-value)</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>F-test of female interactions (p-value)</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Observations</td>
<td>176</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0952</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses. All specifications include wave fixed effects and controls for gender of match treatment. The sample is restricted to participants who received negative feedback in the additional feedback treatment and participants who would have received negative feedback in the limited feedback treatment. Columns 5 and 6 omit wave 3 since it did not contain the attribution question. Even numbered columns control for risk and confidence. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.
statistically significant. This disproportionately changes the entry behavior of high-ability women, even though they received negative feedback primarily due to bad luck, either because they were paired with someone who scored higher, or because they, by chance, happened to be in wave 1. Online Appendix Tables G10, G11, and G12 confirm that the results in Table 6 are qualitatively similar when we include demographic controls, use a logit specification for tournament entry, or use score quartiles instead of bins, respectively.

6. Consequences of the gender asymmetry in attribution: Winners and losers

The results from our experiment highlight the nuanced nature of how gender differences in attribution translate into different outcomes for men and women. Because men are mostly able to attribute negative feedback correctly, feedback benefits them in terms of improving their tournament entry decision. With fewer low-performing men entering in the AF condition, feedback can increase efficiency by making AF tournaments more competitive. However, because high-ability women misattribute negative feedback, fewer high-performing women enter in the AF condition. This effect of feedback, by itself, can reduce efficiency, making AF tournaments less competitive. The combination of these two factors implies an ambiguous effect of additional feedback on overall efficiency, but a clear redistribution of earnings from high-ability women to high-ability men relative to the LF condition.

Panel A of Fig. 6 compares the scores of men and women across the two treatments, conditional on entering the tournament, and suggests that the statistically significant gains by men in the AF condition (p < 0.01) relative to the LF condition may outweigh the modest, statistically insignificant decrease in tournament scores for women (p = 0.86).15 (The gender gap in performance that emerges in the AF condition is significant, p = 0.006, and the difference in the gender gap from the LF to the AF condition is also significant, p = 0.02).

However, tournament entry is endogenous by design. The primary concern about the gender gap in tournament entry is that women lose out on positive returns by not entering tournaments in the first place. Therefore, Panels B and C of Fig. 6 consider the performance in round 2 without conditioning on tournament entry for all high- and low-performing participants, respectively. Panel B shows that the high-performing women in the round 2 task perform better than the high-performing men in the LF condition (p < 0.01). With additional feedback, the high-ability women perform worse (p = 0.09), on average. Low-ability women perform directionally worse than the low-ability men in round 2 in both conditions (statistically significant in the AF condition, p = 0.01), although the difference in the gender gap between conditions is not significant (Panel C). This reduction in women’s performance is an unanticipated negative consequence of noisy feedback worth highlighting.

7. Discussion and directions for future research

This paper describes an online experiment designed to study gender differences in attribution of noisy feedback to luck and ability and explore whether these differences can explain tournament entry patterns. First, we find an asymmetry in the way men and women attribute feedback, particularly when the feedback is negative. Women attribute negative feedback to lack of ability, regardless of whether the feedback is consistent with their original expectation. On the other hand, men attribute negative feedback to lack of ability only when the feedback confirms their pre-existing negative self-evaluation. When men expect a positive outcome, they attribute negative feedback to bad luck.

These gender differences in attribution of negative feedback have real-world consequences in driving the decision to enter a tournament. On average, in our experiment, men are significantly more likely than women to enter competition in the absence of feedback on relative standing. The gender gap is particularly strong when women face a male opponent or when the gender of the opponent is unknown. This is not surprising, because men actually perform significantly better in the MRT task. What is more surprising is that, absent additional feedback, women positively self-select into the tournament based on score, with high-ability women entering the tournament at high rates, while low-ability women stay out. While not something we anticipated ex-ante, this novel result is worth noting, because, contrary to much of the previous findings, high-ability women in our setting do not universally shy away from competing, even though the task is perceived to be (and actually is) male-oriented.

In our experiment, feedback about relative payment eliminates the gender gap in tournament entry, on average. However, unlike other studies that find that feedback improves tournament entry of high-ability women (e.g., Brandts, Groenert, and Rott 2014), in our experiment, additional feedback decreases the rate of tournament entry among low-performing men and has no effect on women, on average. In fact, high-ability women who receive negative feedback significantly decrease their tournament entry relative to comparable women in the limited feedback condition. Gender differences in attribution help explain these results. Recall that women attribute negative feedback to lack of ability, regardless of whether it is consistent with their self-
evaluation. This explains why high-ability women opt out of competition after receiving negative feedback. On the other hand, men react to negative feedback based on their original expectation of ability. Upon receiving a negative signal, men who already had a negative self-evaluation attribute it to low ability (mostly correctly) and opt out of the tournament. But men with a positive self-evaluation attribute this bad news to luck (again mostly correctly) and continue to enter the tournament. We lack sufficient variation in ability among those who received positive feedback to say anything about gender differences in response to positive feedback and leave this to future work.

These observations have potentially important implications for the labor market. In particularly risky environments, where luck is a large component in determining outcomes, such as in financial markets, women receiving negative feedback may benefit from a reminder that luck is a big factor. Otherwise, unlucky female investors, for example, may misattribute losses to low ability rather than luck and may be deterred from making future investments. An investigation of the effect of attribution of feedback on more real-world outcomes, such as investing behavior or educational track choice, and over a longer time horizon, is a fruitful direction for future research.

Because our experimental setting is stylized, the findings should be interpreted with caution. By design, we abstract away from reasons to compete other than to win the tournament, which means that sorting based on actual or perceived ability is the main mechanism for efficient selection. Therefore, our experiment is most directly applicable only to environments where competition serves the sole purpose of succeeding at the particular task. Our results suggest that negative feedback would be interpreted differently by men and women in such environments, potentially exacerbating gender gaps.

We further hypothesize that attribution of feedback can be important in real-world competitions that have benefits above and beyond those directly tied to one’s performance. For example, the act of competing itself may serve as a signal of future performance or other attributes correlated with performance, such as ambition and confidence. Moreover, persistence in challenging or competitive fields despite negative feedback can also lead to an improvement in ability over time, as individuals learn and accumulate human capital. This is particularly relevant to the gender gap in major and career choice (Goldin 2013, Buser, Niederle, and Oosterbeek 2014, Kugler, Tinsley, and Ukhaneva 2017).

Gender differences in attribution of feedback are more likely to play a role here, because negative feedback attributed to bad luck could lead to persistence in the competitive track (men), while negative feedback attributed to lack of ability could lead to dropping the career track (women). Thus, in future work, we plan to investigate the consequences of gender differences in attribution for economic outcomes where dropping out due to perceived low ability may not be the optimal course of action in the long run.

In conclusion, our findings imply that, because of the gender asymmetry in attribution, feedback can lead to positive as well as negative effects on efficiency. On one hand, feedback can increase efficiency, because it reduces inefficient entry of low-performing men into the tournament. On the other hand, surprising negative feedback decreases tournament entry by high-ability women who would have otherwise entered, which can decrease efficiency. Without taking a stand on which of these effects dominate, we note that noisy feedback can exacerbate gender differences in outcomes.

Supplementary materials


References


World Bank (2019). Proportion of seats held by women in national parliaments (%)
https://data.worldbank.org/indicator/SG.GEN.PARL.ZS.