LUCK OR SKILL: HOW WOMEN AND MEN ATTRIBUTE SUCCESSES AND FAILURES[*]

G. Kartini Shastry[†] Olga Shurchkov[‡] Lingjun "Lotus" Xia[§]

Wellesley College Wellesley College Wellesley College

First Draft: August 2018

Current Draft: August 2018

Abstract:

Gender differences in preferences, such as risk aversion, are often used to explain gender gaps in wages and other economic outcomes. The key factor in linking preferences to outcomes is decision-making, yet the mechanisms behind this link remain unknown. One possibility is that women and men react differently to feedback. We design an online controlled experiment where some subjects are given feedback before deciding whether or not to enter a tournament. We find that, even in the absence of feedback, women know when to enter, with low-performers staying out of competition and high-performers entering. On the other hand, without feedback, men do not sort based on performance: low-performing men are significantly more likely to enter than comparably low-performing women. Feedback erases the gender gap in tournament entry by correcting the suboptimal entry pattern of low-performing men. Furthermore, women and men react to feedback differently. Women attribute negative feedback to lack of ability, whether or not the information is consistent with their prior self-evaluation. Men, on the other hand, attribute negative feedback to bad luck when the information contradicts their previously confident self-evaluation. Men attribute negative feedback to lack of ability only when the information confirms their previously below-average self-evaluation. Finally, we find that women are less likely than men to take credit for positive feedback.

Key Words: gender differences, competition, attribution, feedback, economic experiments

JEL Classifications: C90, J16, J71

---

[*] We wish to thank Kristin Butcher, Catherine Eckel, Antonio Filippin, Eric Hilt, Casey Rothschild, and Jeremy Wilmer for helpful discussions and feedback, as well as the participants of the Wellesley College Economics Research Seminar and the University of Tampere Institutions in Context Workshop on Gender Equality and Policy. Xia gratefully acknowledges the Jerome A. Schiff Fellowship for financial support. Shastry and Shurchkov gratefully acknowledge financial support from Wellesley College Faculty Awards. All remaining errors are our own.
[†] Department of Economics, Wellesley College, 106 Central St., Wellesley, MA, USA, gshastry@wellesley.edu
[‡] Corresponding author. Department of Economics, Wellesley College, 106 Central St., Wellesley, MA, USA, olga.shurchkov@wellesley.edu
[§] Wellesley College, 106 Central St., Wellesley, MA, USA, lxia@wellesley.edu.

I. Introduction

Gender gaps in economic outcomes are pervasive and persistent. Only 5 percent of Fortune 500 companies have female CEOs, with women occupying a meager 26.5 percent of executive and senior-level managerial positions (Catalyst 2018). A simple comparison of annual wages for full-time working women and men reveals a gender wage gap of about 79 percent (Blau and Kahn 2017). Leadership gaps are similarly prevalent in politics. In the US, women hold just 19 percent of seats in the House of Representatives and 22 percent in the Senate. The problem is not restricted to the US: in Canada, for example, just over one-quarter of members of the House of Commons are women (Catalyst 2018).

Explanations for these gender gaps can be sorted into two broad categories. First, women might be discriminated against directly due to their gender 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) show that about 50% of the gap is due to occupational choice, while 38% can be attributed to pure discrimination. However, choices could be made in anticipation of discrimination and may not simply reflect different preferences.

In this paper, we focus on gender differences in behavioral traits that may serve as underlying mechanisms behind differential sorting into careers and opportunities (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. Researchers have set out to investigate ways in which women may be "nudged" to be more competitive. The provision of feedback has been a natural intervention that has nonetheless yielded mixed results (Ertac and Szentes 2011; Brandt, Groenert, and Rott 2014). Our experiment contributes to the literature by shedding light on the effect of feedback about one's standing relative to others, when the feedback is determined partly by individual ability and partly by luck.

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 and partly on whether their performance is higher or lower than that of a randomly chosen participant. They are then randomized into two main treatment conditions: control with no feedback and treatment with feedback about how their individual payment relates to the average payment of a comparable group of participants. In the second round, subjects perform the same task again, but first choose between a piece-rate scheme and a tournament-based payment scheme. We replicate the robust finding in the literature (NV) of the gender gap in competitiveness: in the no-feedback condition, men enter the tournament significantly more frequently than women, on average. However, our results challenge the notion that women should compete more in such tournaments: we find that women know when to enter, even without feedback, with low-performers staying out of competition and high-performers entering. On the other hand, without feedback, men do not sort based on performance: low-performing men are as likely to enter as high-performing men. Positive sorting based on ability by women is consistent with similar findings that women act strategically in different contexts (see, for example, Exley, Niederle, and Vesterlund 2016 for bargaining).

Despite performing a task perceived to favor men, there is no gender gap in scores when we look at actual performance by those who entered the tournament in the no feedback condition. This is explained by the fact that high-performing women are competing alongside high- and low-performing men. Our first contribution is to find that feedback erases the gender gap in tournament entry, but not because it changes the behavior of women. In fact, women continue to sort into tournament appropriately. Feedback eliminates the tournament entry gap by correcting the suboptimal entry pattern of low-performing men who now choose not to enter competition. Consequently, feedback generates a gender gap in performance: on average, men who enter the tournament in the feedback condition outperform the women who enter the tournament.

Our second finding arises from the analysis of gender differences in assigning responsibility for particular outcomes, the so-called attribution of feedback. It sheds new light on potential reasons behind the persistent gender gaps in choosing to stay in certain educational or career tracks. In particular, if women are quick to blame themselves for negative feedback, while men are more likely to blame bad luck, then we would expect to see more men staying on competitive tracks and more women dropping out. In our experiment, subjects receive information about their relative payment, and are subsequently asked to assess the extent to which they believe the outcome arose due to luck or ability (bad/low if the payment is below average or good/high if the payment is above average). We find that women consistently attribute negative feedback to low ability, regardless of whether the feedback is consistent with their original self-evaluation. However, men attribute negative feedback to low ability only when the feedback confirms a pre-existing negative self-evaluation. When men are confident about their performance being above average, but then receive negative feedback, they attribute the negative feedback to bad luck.

This paper relates to several literatures. First, a large body of experimental work has tested the extent to which gender differences in behavioral traits, such as risk aversion, competitiveness, or social preferences, contribute to the gender gap in labor market outcomes (see Shurchkov and Eckel 2018 for a survey of the literature). Specifically, 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. As noted above, several papers have examined whether feedback affects the gender gap in tournament entry. Ertac and Szentes (2011) find that information about the performance of others eliminates the gender gap in tournament entry. However, the gender gap in competitiveness persists after participants receive advice about tournament entry from a better informed person (Brandt, Groenert, and Rott 2014). Our results reveal that feedback on something as simple as relative payment can eliminate the tournament entry gender gap.

This paper also contributes to the literature on how feedback differentially affects beliefs by gender. Mobius et al. (2011) found that men who receive feedback are more likely to revise their beliefs about performance as compared to women who receive similar feedback. Berlin and Dargnies (2016), however, found the opposite result: women react more strongly than men to feedback, becoming more confident after receiving positive feedback and less confident after negative feedback. The seeming inconsistency can be reconciled by the difference in feedback accuracy across the two studies. Like in Berlin and Dargnies (2016), feedback in our experiment is 100% accurate, but the novel feature of our design is that the signal about performance is noisy, resembling more closely the real world, where feedback is determined by both ability and luck.

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 some external forces, such as luck, in order to maintain self-esteem (Miller and Ross, 1975; Mezulis et al. 2004; see Eil and Rao 2011 for application in economics). Our study explores self-attribution in the context of reactions to feedback by asking the explicit attribution question and shedding light on how responses vary by gender. Our findings of gender differences in attribution of feedback provide a novel potential explanation for recent findings of gender differences in tournament entry even after receiving feedback. For example, Buser and Yuan (forthcoming) show that women are less likely than men to compete after losing in an earlier round.

The rest of the paper is organized as follows. Section II describes the experimental design and provides a first look at the experimental data. Section III summarizes observed gender differences in preliminary outcomes, such as ability in the task, and behavioral traits such as confidence and risk attitudes. Section IV presents our main findings on the effect of feedback on selection into competition. Section V reports gender differences in attribution of feedback. Section VI concludes with some implications of the results and discusses potential future research.

II. The Online Experiment

Our controlled online survey experiment establishes an environment where we can observe the effect of feedback on sorting into competition and on how men and women differ in their attribution of that feedback.

The experiment was programmed in Qualtrics and conducted using the Amazon Mechanical Turk (AMT) platform. 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. We used randomly placed attention checking questions in order to ensure full attention. A full set of screen shots with the consent form and experimental instructions is available in the online Appendix (Section A).

II.A. Design of Section 1: Measuring Attribution across Treatments

In our experiment, participants began by completing Section 1 of problem solving (see Section II.B for the detailed description of the task). Participants had exactly 2.5 minutes to complete as many questions as possible out of 8 (the time ran out automatically). After completing the task, participants estimated their scores for that section of problem-solving. The resulting value measures each subject's *score confidence*.

Importantly, participants did not know their actual score from problem-solving at any point in the experiment. Instead, they received information about their payment. In particular, each participant learned 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.[1] If their score was higher than or equal to that of their random match's, they got 20 cents for each correct answer. However, if their score was lower than that of their random match's, they got only 15 cents for each correct answer. The participants then saw their payment (in US$). The exact wording of the payment information varied according to random assignment of participants into two treatments: match gender known condition or match gender unknown condition.

[UG] Match with unknown gender

In this treatment group, participants saw their payment without learning anything about the gender of their random match. In particular, they viewed the following on-screen text:

"*Our matching process has randomly matched you with a participant from the other group. Your score has been compared with his/hers, and your payment is shown below*."

We used a gender noncommittal pronoun "his/hers" to express that the random match might be of either gender.

[KG] Match with known gender

In this treatment group, the on-screen payment information text subtly revealed the 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.[2]

"*Our matching process has randomly matched you with a female (male) participant from the other group. Your score has been compared with hers (his), and your payment is shown below*."

The choice of "female" (hers) and "male" (his) depends on the actual gender of the random match. The gender specification is made implicit but repeated twice ("female participant," and

[1] 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 appendices B and C for detailed information on pilot waves 1 and 2.
[2] Our conclusions are qualitatively robust to restricting the sample to participants who were randomized into the unknown gender treatment.

“her,” for example) to prevent participants from guessing the purpose of the experiment, and yet to ensure they pick up the cue.

Upon receiving payment information, subjects in both treatments 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 appendices B and 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*.”

Next, participants were further randomized into three treatment conditions based on the amount of feedback they would receive before answering a question regarding how they attribute their payment outcome. To measure *attribution*, we asked all participants to use a slider in order to indicate the relative importance of luck, as opposed to their own ability, in determining their payment outcome (see Figure 1).

[FIGURE 1 ABOUT HERE]

[NF] No feedback condition

In this condition, participants did not receive any additional information before moving on to the attribution question. Thus, these participants assessed the contribution of luck and skill to their payment purely based on their perception of relative payment – their self-evaluation. For example, a participant with a positive self-evaluation assessed to what extent her presumed above-average payment was due to her own high ability, and to what extent it was due to good luck of being matched with a relatively weak participant from the other group. On the other hand, a participant with a negative self-evaluation estimated to what extent her presumed below-average payment was due to her own lack of ability, and to what extent it was due to bad luck of being matched with a relatively strong player from the other group.

[FF] Forced feedback condition

In this condition, participants learned whether their payment was *actually* above or below average. Therefore, there was no uncertainly about relative performance, and these participants assessed the contribution of luck and skill to their payment based on actual feedback, rather than a self-evaluation.[3] For example, a participant receiving positive feedback (payment above average) assessed to what extent her above-average payment was due to her own high ability, and to what extent it was due to good luck of being matched with a relatively weak player from the other group. On the other hand, a participant receiving negative feedback (payment below average) estimated to what extent her below-average payment was due to her own lack of ability, and to what extent it was due to bad luck of being matched with a relatively strong player from the other group.

[OF] Optional feedback condition

In this condition, participants had an opportunity to give up some of their payment in exchange for feedback on relative payment. Specifically, participants indicated their willingness to pay to receive feedback from a list of options ranging from 5 cents to 25 cents. The computer then randomly drew a price from the list. If the participant indicated that she was willing to pay the randomly drawn price, then the price was subtracted from her payment, and the rest of the design was as in the FF condition. If the participant indicated that she was not willing to pay, then she did not receive any feedback, and the rest of the design was as in the NF condition.

Table 1 reviews the 2x3 factorial design in our experiment and provides the total number of subjects in each treatment condition.

[TABLE 1 ABOUT HERE]

II.B. THE PROBLEM SOLVING TASK

The choice of task in our experiment was based on three criteria. First, we looked for a task that cannot be easily "cracked" by the AMT subjects who, unlike lab subjects, have access to the internet and calculators. Second, we preferred a skill-based task to a menial task, such as the slider task (Gill and Prowse 2012) because of the more natural applications to real-world contexts where gender gaps are the greatest. Finally, our ideal task is either gender-neutral or stereotyped to suit men, although in future work we plan to broaden our study to consider a stereotypically female task.[4]

[3] Our analysis takes into account the fact that attribution is framed differently depending on whether a participant received feedback.

[4] Shurchkov (2012) points out that gender gaps in competitiveness are particularly pronounced in tasks which are perceived to favor men, such as the task we use in this study. In a separate pilot study, we experimented with a stereotypically female-favoring task, namely, the anagram task. However, we found that in many cases participants were able to use online search engines to achieve perfect scores.

In order to find a suitable task to meet the above criteria, we conducted two waves of pilot surveys that tested for gender differences in performance, confidence, and gender perceptions of three potential candidate tasks: Mental Rotation Task (MRT); "find the median" task; and a "pattern" task (also known as MPT or the matrix test). Detailed information on the three tasks, pilot survey protocol, and results can be found in the online appendix, sections B and C.

Based on our pilot results, we settled on using the MRT in our main experiment. It is a test in which 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. In our experiment, subjects received one point for each correct answer and zero points for each wrong or blank answer. Figure 2 shows a sample problem. [5]

[FIGURE 2 ABOUT HERE]

Before completing the first section of problem-solving, each participant was given instructions for solving an MRT problem, as well as a practice example. They could not advance to the real problem-solving section until they correctly solved the practice question.

Within each problem-solving section, 8 MRT questions were presented on a single screen. The order of the questions was randomized for each participant.

II.C. Section 2: Tournament Entry and Questionnaire

In Section 2 of the experiment, subjects once again had 2.5 minutes to solve as many MRT problems as possible. Prior to problem-solving, participants had to choose the type of payment scheme they wished to apply to subsequent performance. The two choices were:

- **Piece rate**: Participant gets 17.5 cents for each correct answer, regardless of anyone else's score.
- **Tournament**: We randomly match the participant with another participant from pilot wave 2. If the score is higher than the score of the random match, then the participant earns 25

[5] The MRT questions we used in our experiment were slightly different from the original MRT (Vandenberg and Kuse 1978). In the original MRT, there are four choices for each target shape. Exactly two of the choices are correct. Participants get 1 point for each correct choice and lose 1 point for each wrong choice. In order to reduce the difficulty level, we took out one of the correct choices for each target shape, and removed the penalty for incorrect choices.

cents for each correct answer. If the score is lower than the score of the match, then the participant only gets 10 cents for each correct answer.

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 much 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. In fact, entering the competition in itself may be a positive signal of confidence that in the real world might be rewarded rather than punished. In Section VI, we will further discuss the applications of this and other aspects of our unique design to real world environments where gender differences are particularly pronounced.

At the end of the experiment, each participant filled out a short questionnaire, which can be found in the online appendix A. First, we elicited participants' risk preferences by asking them to self-report, on a scale of 1 to 10, how willing they are to take risks. Research has shown that the self-reported measure correlates significantly with experimental outcomes for risk-aversion and is widely accepted in the literature (Kagel and Roth 2016).

In order to confirm that our main experimental subjects hold similar gender stereotypes as the pilot subjects, the questionnaire asked the same question about gender perceptions associated with MRT. Finally, the questionnaire also collected information on demographic characteristics, including age, gender, race, level of education, and income.

After participants finished the experiment, they were paid a base payment of $0.50 for completion. The final payment, including their bonus payment earned in the two problem-solving sections, was transferred to their account within seven days. Including the base payment, the average payment was $2.1. The maximum payment was $4.1. The average duration of the experiment was about 11 minutes.

II.D. A First Look at the Data

We conducted three waves of data collection in the main experiment. In the first wave, we administered 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 focused on the forced feedback (FF) condition which we identified at the outset as our primary treatment of interest, collecting 88 additional valid responses. We also corrected a small coding error that affected the FF condition but did not preclude us from using the data from wave 1.[6] In the third wave, we collected more data in the NF and FF treatment conditions, but decided to exclude the optional feedback condition. Very few participants were interested in receiving feedback, which made it difficult to compare those who sorted into receiving feedback to those who did not. We therefore omit the optional feedback condition data from all subsequent analysis. In the third wave, we also omitted the attribution question, in light of the potential concern that attribution might prime tournament entry. A total of 198 valid responses were collected in the third wave.

Table 2 reports the summary statistics of demographic characteristics and performs balance tests between treatment groups based on demographics. The educational background of study participants is fairly similar to the national average in the United States, though skewing slightly towards being more educated (Ryan and Siebens 2016). The racial makeup in our study skews toward Whites and away from African-Americans and other minorities (US Census). The income makeup in our study is similar to the national distribution (US Census). Only 9 out of 94 comparisons across treatments are significant at the 0.05 level. All the procedures are randomized in the survey, but we conduct robustness checks and find that our results are robust to the inclusion of demographic controls.

[TABLE 2 ABOUT HERE]

In terms of the gender difference in demographic variables, women are slightly older than men in our sample (mean age for men is 37 and mean age for women is 40; $p < 0.0001$) and men are less likely than women to have attended but not graduated from college (19 percent of men and 27 percent of women; $p = 0.02$). Men and women are statistically identical to one another according to all other demographic characteristics (tests are available upon request).

[6] A coding error in Wave 1 of the experiment resulted in some participants receiving negative feedback when they should have received positive feedback. The error did not affect anything other than the feedback received by some subjects, so we were able to use the data in our analysis. Our results are similar when we drop the 45 observations affected by this error, although we lose some statistical power.

III. Gender Differences in Ability and Behavioral Traits

Table 3 reports summary statistics of key experimental variables by gender. All of these variables are from the Section 1 task, collected before subjects were randomized into feedback treatment conditions, except for self-reported risk preference. We pool the data across the three waves and across treatments.[7] The order of variables reported in the table follows the chronological order of the design of the experiment.

[TABLE 3 ABOUT HERE]

We find that men outperform women in the MRT task and consequently earn a higher bonus in Section 1. Indeed, there is a consensus in psychology that MRT consistently elicits gender differences in performance (Masters and Sanders 1993).

Women in our experiment are significantly less confident than men, both in terms of their prediction of Section 1 score (score confidence) and in terms of whether they believe their payment to be above or below average (confidence about relative payment).[8] A simple comparison of actual and predicted score for men and women reveals that it is women who are significantly under-confident in our experiment (mean confidence of 3.53 relative to actual average score of 3.79, t-test p-value of 0.039). Men also underestimate their score, on average, but the error is not statistically significant (mean confidence of 3.86 relative to actual average score of 4.05, t-test p-value of 0.143). Figure 3 plots the relationship between participants' expected and actual scores for the entire distribution of ability. We make three observations: 1) Men are systematically more score-confident ($p < 0.05$) than women, conditional on getting the same score. 2) Men with median performance (solved 4 out of 8 questions) on average correctly estimate their scores. Women at the median, on the other hand, underestimate their score. 3) Participants of both genders with higher-than-median performance tend to underestimate their score. Participants of both genders

[7] We omit the optional feedback condition treatment arm in this table for consistency with the rest of our analysis, even though most of these outcomes are collected before the treatment condition is revealed. The results persist if we include all observations.

[8] Our experiment also replicates the 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). Filippin and Gioia (2018) find that males become significantly more risk averse after losing a tournament than after randomly earning the same low payoff. Since the participants self-select into our tournament, we are unable to test this hypothesis in our data.

with lower-than-median performance tend to overestimate their score.[9] Gender differences in confidence are a robust finding in the experimental literature (see for example, Beyer 1990), although men are typically found to be universally overconfident in lab experiments (see for example, NV).

[FIGURE 3 ABOUT HERE]

Table 4 uses ordinary least squares to confirm that the gender gap in both confidence measures decreases somewhat but remains significant if we control for actual performance (Columns 2 and 4). The finding is robust to controlling for wave fixed effects and including demographic controls (estimates are available upon request).

[TABLE 4 ABOUT HERE]

IV. Effects of Feedback on Tournament Entry and Sorting

We begin the discussion of our main results with a look at tournament entry patterns in our data. In their seminal paper, NV show that women shy away from competition. As seen before (Shurchkov 2012), the environment turns out to be important. The NV study focuses on a mathematical task (different from ours, but similar from the gender-stereotype perspective), in a setting where women are asked to compete in mixed gender groups (two men and two women), and where feedback on relative performance is not available. The conditions most consistent with NV's setting in our experiment are either the mixed gender treatment with no feedback or the instances of facing a male opponent with no feedback. Figure 4 shows that the gender gap in tournament entry is significant and economically large under these conditions, replicating the NV result (dark grey bars, showing the tournament entry share of 0.47 for men and 0.29 for women, p-value of 0.022). Note that the result replicates even though tournament entry in our experiment is less risky than in NV and other similar studies, with the loser still receiving some payment.

[FIGURE 4 ABOUT HERE]

[9] For high-performing women: Mean(actual score) = 5.74; Mean(expected score) = 4.48; $p < 0.01$. For high-performing men: Mean(actual score) = 6.05; Mean(expected score) = 5.11; $p < 0.01$. For low-performing women: Mean(actual score) = 2.23; Mean(expected score) = 3.02; $p < 0.01$. For low-performing men: Mean(actual score) = 2.35; Mean(expected score) = 3.44; $p < 0.01$).

Next, we highlight the two institutional changes that help to eliminate the gap in competitiveness in our setting. First, we continue with the no-feedback environment, and observe that facing a female match (i.e., opponent) significantly increases selection into tournament by women, 55 percent of whom compete in this condition. On the other hand, the knowledge that the opponent is female actually decreases entry by men to 33 percent, although the reduction in competitiveness for men is not statistically significant. The small sample of female-to-female matches in the no feedback condition warrants caution when interpreting the results, and indeed the reversal of the gender gap in that condition is not statistically significant at conventional levels (the comparison of 55 percent entry for women to 33 percent entry for men produces a p-value of 0.104).

Figure 4 also demonstrates that feedback eliminates the gender gap in tournament entry. Ertac and Szentes (2011) find a similar headline result, but in their study, which uses a winner-take-all tournament scheme like NV, performance feedback increases entry by high-performing women without a significant change to the behavior of high-performing men or to the behavior of low-performers of either gender. The right panel of Figure 4 demonstrates that our result is quite different: feedback reduces tournament entry by men, without an effect on tournament entry by women. This observation leads to a natural follow-up investigation of whether feedback might be correcting some inefficiency in sorting that exists in the absence of feedback.

Figure 5 plots the relationship between score in Section 1 and the probability of tournament entry for men and women in the no feedback condition and in the forced feedback condition. In the absence of feedback, women are generally more likely to enter tournament as their score in Section 1 increases. However, for men, there is a slight U-shaped relationship, if any, between tournament entry and score: low-performing men are equally likely to enter competition as high-performing men. When feedback on relative payment is available, men react to it, while women do not. In particular, feedback on relative payment corrects the suboptimal entry pattern of low-performing men, but does not substantially alter the behavior of high or low-performing women.

[FIGURE 5 ABOUT HERE]

Table 5 verifies that the pattern of tournament selection by gender holds when we apply a linear probability model.[10] Note that we control for both genders interacted with section 1 score, leaving out the main effect, for ease of interpretation. The p-value at the bottom of the table indicates whether the impact of section 1 score on tournament entry differs for men and women. In the no feedback condition, women are marginally less likely to enter the tournament on average (Column 1, row 1). Scoring one point higher in Section 1 is associated with an approximately seven 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. Column 2 shows that risk preferences and confidence do not explain the pattern of tournament entry (even when we interact them with gender; results available upon request). These results are consistent with the finding of women acting strategically in other contexts. For example, Exley, Niederle, and Vesterlund (2016) observe that women positively select into negotiations, whereas men do not.

[TABLE 5 ABOUT HERE]

Columns 3 and 4 of Table 5 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 4 when controls are included (the p-value is 0.104 between Columns 1 and 3 without controls). As a result, feedback erases the gender gap in tournament entry; the gender gap is significantly different in the forced feedback relative to the no feedback condition. The finding is robust to the inclusion of risk and confidence as additional explanatory factors.

Note that, while feedback erases one type of gender gap in our experiment, it generates another. In particular, among those who do not receive any feedback, women and men who enter the tournament perform equally well in Section 2 of MRT (mean score of 4.02 for men and 4.36 for women; p-value of 0.497). Recall that the MRT favors men on average, but a large share of low-performing men who did not receive feedback entering the tournament causes the average score in the tournament to be relatively low. On the other hand, feedback causes low-performing men to

[10] Logistic specifications produce similar results and estimates are available in the online appendix E.

drop out of competition, which leads to a significant gender gap in performance in the tournament in the feedback condition (mean score of 5.40 for men and 4.67 for women; p-value of 0.057).

V. Gender Differences in Attribution of Feedback

In the previous section, we showed that access to information changes the behavior of subjects (men). We next investigate the role of information by asking how different types of feedback is perceived by the two genders. In particular, there can be 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 negative self-evaluation (i.e., negative surprise). In order to shed light on how subjects react to these types of feedback, we restrict our attention to the forced feedback condition, where subjects assess the attribution of actual relative payment information to luck and to own ability. We also restrict our attention to the participants in waves 1 and 2, as the attribution question was omitted from wave 3. The attribution measure ranges from 0 (attributing payment outcome feedback completely to luck) to 100 (attributing payment outcome feedback completely to self).

Across all feedback conditions, men and women do not significantly differ in their attribution, with men attributing 66.1 percent of the outcome to their own ability and women attributing 63.9 percent of the outcome to their own ability, on average. However, attribution varies with the type of feedback subjects receive. Table 6 reports average attribution by gender and type of feedback condition.

[TABLE 6 ABOUT HERE]

The top two rows of Table 6 demonstrate that men and women who hold a negative self-evaluation (i.e., originally believed they had a below-average payment) act similarly in terms of attributing feedback. Both men and women who receive reinforcing negative feedback of their payment actually being below average attribute this adverse outcome to their own relatively low ability more than to luck (about 70 percent to ability and 30 percent to luck). Both men and women who are surprised to learn that their payment was actually above average still attribute much of that outcome to own ability, albeit less than when they received reinforcing negative feedback.

The bottom two rows of Table 6 reveal that men and women react to feedback substantially differently when they originally hold a positive self-evaluation (i.e., originally believed they had an above-average payment). Men who receive negative feedback are significantly more likely than women to attribute it to bad luck. However, when feedback reinforces the self-evaluation as positive, men are significantly more likely than women to take credit for it as being due to their high ability.

Table 7 presents the estimates from a linear regression model that provides further insight into the gender gap in attribution of feedback. The omitted category in all specifications is a man who receives positive feedback.

[TABLE 7 ABOUT HERE]

In Column (1), among those who hold a negative view of their relative payment – a below-average self-evaluation – both men and women are more likely to attribute reinforcing negative feedback to own ability. The inclusion of additional controls for score in Section 1, score confidence, and risk attitudes in Column (2) only slightly decreases the magnitude of the estimates, so that the difference for women is no longer significant, but the attribution by men to self remains.

In Column (3), among those who hold a confident view of their relative payment – an above-average self-evaluation – attribution of feedback greatly varies by gender. First, note that women are significantly less likely to attribute reinforcing positive feedback to their own high ability, as compared to men who also receive positive feedback (the 13.65-point difference is significant and large in magnitude). Second, men who receive surprising negative feedback are significantly more likely to attribute the bad news to bad luck. On the other hand, women who receive the same surprising negative feedback are significantly more likely to blame themselves, attributing it to low ability. The gender gap in reaction to surprising negative feedback is large and statistically significant (F-test p-value < 0.0001). The results are robust to the inclusion of controls for score in Section 1, score confidence, and risk attitudes in Column (4).

Our results are broadly similar when we compare attribution of subjects in the no feedback condition to those in the forced feedback condition. Note that, whether or not a participant receives feedback determines whether she is attributing her perceived payment or her actual payment, which potentially complicates the interpretation of the effects. The full analysis can be found in the online appendix D.

To summarize, we find that women consistently attribute negative feedback to lack of ability, regardless of whether the feedback is consistent with their original self-evaluation. However, 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, so that negative feedback undermines their positive self-evaluation, they attribute this feedback to bad luck.[11]

VI. Discussion and Directions for Future Research

This paper describes an online experiment designed to uncover gender differences in reaction to feedback, both in terms of its effect on sorting into tournaments and in terms of the psychological perception about one's responsibility for the outcome.

In regards to tournament entry, we find that men are significantly more likely than women to enter competition in the absence of feedback. The gender gap is particularly strong when women face a male opponent or when the gender of the opponent is unknown. On average, feedback eliminates the observed gender gap in tournament entry. A closer look at sorting patterns reveals that, in the absence of feedback, women positively self-select into tournament based on score, whereas men do not – a finding that is consistent with Exley, Niederle, and Vesterlund (2016) who tell a similar story in the context of bargaining. In our experiment, subsequent feedback about relative payment decreases the rate of tournament entry among low-performing men, erasing the gender gap in competitiveness, but generating a gender gap in performance under competition. The implication of these results is that it may not be optimal to encourage women to enter such competitions, especially when the competition favors men.

[11] One interesting connection between our two sets of results is in the sample of people driving each result. The gender difference in tournament entry in response to feedback was driven by changes in the behavior of low-performing men. The gender difference in attribution of feedback was also driven by changes in the behavior of men, but in this case men with positive self-evaluation. Further exploring these differences, we find that while the attribution results are similar among low and high performing participants, the tournament entry results are driven by participants with negative self-evaluations. This is internally consistent. Our second set of findings shows that when men with negative self-evaluations received negative feedback, they attributed it to lack of ability. Those who received negative feedback are more likely to be the low-performing men, who according to our first result respond to the feedback by opting out of the tournament.

In regards to attribution of feedback, we find that women attribute negative feedback to lack of ability, regardless of whether it is consistent with their self-evaluation. Women are also more likely than men to attribute surprising positive feedback to luck as compared to men who similarly receive surprising good news. On the other hand, men attribute negative feedback to lack of ability only when they hold a negative self-evaluation initially. When surprising negative feedback undermines their positive self-evaluation, men attribute this bad news to luck. 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 traders, 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, and over a longer time horizon, is a fruitful direction for future research.

A consequence of feedback eliminating the gender gap in tournament entry is that it makes it difficult to study the extent to which gender differences in attribution of feedback explain gender differences in tournament entry. In the stylized context of the experiment, where the benefits of entering the tournament are completely driven by performance, sorting based on actual or perceived ability is the main mechanism for efficient selection. We find that women are particularly effective at sorting based on performance. In this sense, our experiment is most directly applicable to environments where competition only serves one purpose, and that is to succeed at the particular task, such as attaining publication at a highly selective journal. In this setting, the act of submitting a paper to a high-ranking journal does not produce any further benefits other than trying to publish there. Our experimental results suggest that feedback alone can increase efficiency in such environments.

However, we 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 on economic outcomes where sorting out due to perceived low ability may not be the optimal course of action in the long run.

References

Berlin, Noémi, and Marie-Pierre Dargnies, 2016. "Gender Differences in Reactions to Feedback and Willingness to Compete," *Journal of Economic Behavior and Organization* 130: 320-336.

Beyer, Sylvia, 1990. "Gender differences in the accuracy of self-evaluations of performance," *Journal of Personality and Social Psychology* 59 (5): 960.

Blau, Francine D., and Lawrence M. Kahn, 2017. "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature,* 55(3): 789-865.

Brandts, Jordi, Valeska Groenert, and Christina Rott, 2014. "The Impact of Advice on Women's and Men's Selection into Competition," *Management Science* 61(5): 1018-1035.

Buser, Thomas, Muriel Niederle and Hessel Oosterbeek, 2014. "Gender, Competitiveness, and Career Choices," *Quarterly Journal of Economics,* 129(3): 1409-1447.

Buser, Thomas and Huaiping Yuan, forthcoming. "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad," *American Economic Journal: Applied Economics*.

Catalyst, 2018. "Knowledge Center: Women in S&P 500 Companies," http://www.catalyst.org.

——, 2018. "Knowledge Center: Women in Government," http://www.catalyst.org.

Eil, David, and Justin M. Rao, 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics* 3 (2): 114-138.

Ertac, Seda, and Balazs Szentes, 2011. "The Effect of Performance Feedback on Gender Differences in Competitiveness: Experimental Evidence, Koç University-TUSIAD Economic Research Forum Working Papers 1104.

Exley, Christine L., Muriel Niederle, and Lise Vesterlund, 2016. "Knowing When to Ask: The Cost of Leaning in," NBER Working Paper w22961, National Bureau of Economic Research.

Filippin, Antonio, and Francesca Gioia, 2018. “Competition and Subsequent Risk-taking Behaviour: Heterogeneity across Gender and Outcomes,” *Journal of Behavioral and Experimental Economics*, 75: 84-94.

Germine, Laura, Ken Nakayama, Bradley C. Duchaine, Christopher F. Chabris, Garga Chatterjee, and Jeremy B. Wilmer, 2012. “Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments,” *Psychonomic Bulletin & Review* 19 (5): 847-857.

Gill, David, and Victoria Prowse, 2012. “A Structural Analysis of Disappointment Aversion in a Real Effort Competition,” *American Economic Review*, 102: 469–503.

Goldin, Claudia, 2014. “A Grand Gender Convergence: Its Last Chapter,” *American Economic Review* 104 (4): 1091-1119.

Goldin, Claudia. “Can ‘Yellen Effect’ Attract Young Women to Economics?” Bloomberg View, Oct 14, 2013. https://www.bloomberg.com/view/articles/2013-10-14/can-yellen-effect-attract-young-women-to-economics, Accessed Mar 1, 2017.

Kagel, John H., and Alvin E. Roth, eds. 2016. *The Handbook of Experimental Economics, Volume 2*. (Princeton University Press, Princeton, NJ).

Kugler, Adriana D., Catherine H. Tinsley, Olga Ukhaneva, 2017. “Choice of Majors: Are Women Really Different from Men,” Working Paper 23735, National Bureau of Economic Research.

Masters, Mary. S. and Barbara Sanders, 1993. “Is the Gender Difference in Mental Rotation Disappearing?” *Behavior Genetics* 23 (4): 337.

Mezulis, Amy H., Lyn Y. Abramson, Janet S. Hyde, and Benjamin L. Hankin, 2004. “Is There a Universal Positivity Bias in Attributions? A Meta-Analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias, *Psychological Bulletin* 130 (5): 711.

Miller, Dale T., and Michael Ross, 1975. “Self-serving Biases in the Attribution of Causality: Fact or Fiction?” *Psychological Bulletin* 82 (2): 213.

Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat, 2011. “Managing Self-Confidence: Theory and Experimental Evidence.” Working Paper 17014, National Bureau of Economic Research.

Niederle, Muriel and Lise Vesterlund, 2007. “Do Women Shy Away from Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics* 122 (3): 1067-1101.

Niederle, Muriel, 2016. “Gender,” Chapter 8 in *The Handbook of Experimental Economics, Volume 2*, John Kagel and Alvin E. Roth, eds. (Princeton University Press, Princeton, NJ).

Paolacci, Gabriele, Jesse Chandler, and Panagiotis Ipeirotis, 2010. “Running Experiments on Amazon Mechanical Turk,” *Judgment and Decision Making* 5 (5): 411-419.

Rand, David, 2012. “The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments,” *Journal of Theoretical Biology* 299 (21): 172-179.

Ryan, Camille and Julie Siebens, 2016. “Educational Attainment in the United States: 2015,” (PDF). U.S. Census Bureau. Retrieved December 22, 2017.

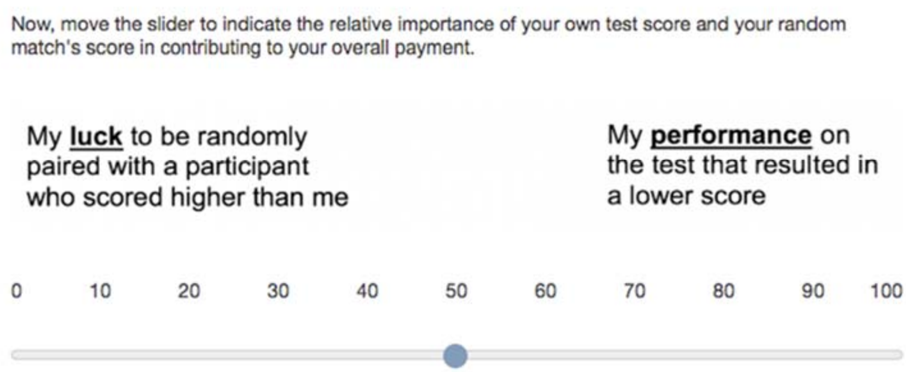
Sarsons, Heather, 2017. “Gender Differences in Recognition for Group Work: Gender Differences in Academia,” *American Economic Review: Papers and Proceedings*, 107 (5), 141-145.

Shurchkov, Olga, 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.

Shurchkov, Olga, and Catherine C. Eckel, 2018. “Gender Differences in Behavioral Traits and Labor Market Outcomes,” in *The Oxford Handbook of Women and the Economy*, Susan L. Averett, Laura M. Argys, and Saul D. Hoffman, eds. (Oxford University Press, New York, NY.)

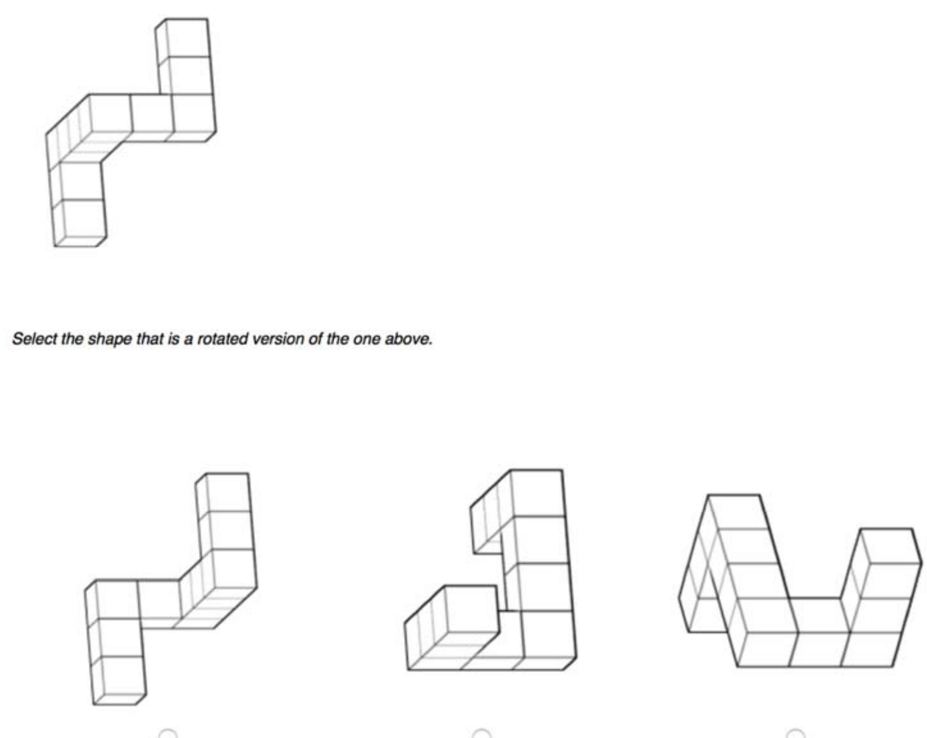
FIGURES

Figure 1: The Attribution Question for a Participant with Negative Information



Notes: The measure of attribution is scaled from 0 to 100. A value of 100 means attributing one's payment entirely to one's own performance on the test; a value of 0 means attributing one's payment entirely to luck of being paired with a specific match.

Figure 2: Sample MRT Problem



Note: The correct answer is the third choice.

Figure 3: The Relation between Expected Score (Score Confidence) and Actual Score in Section 1, by Gender

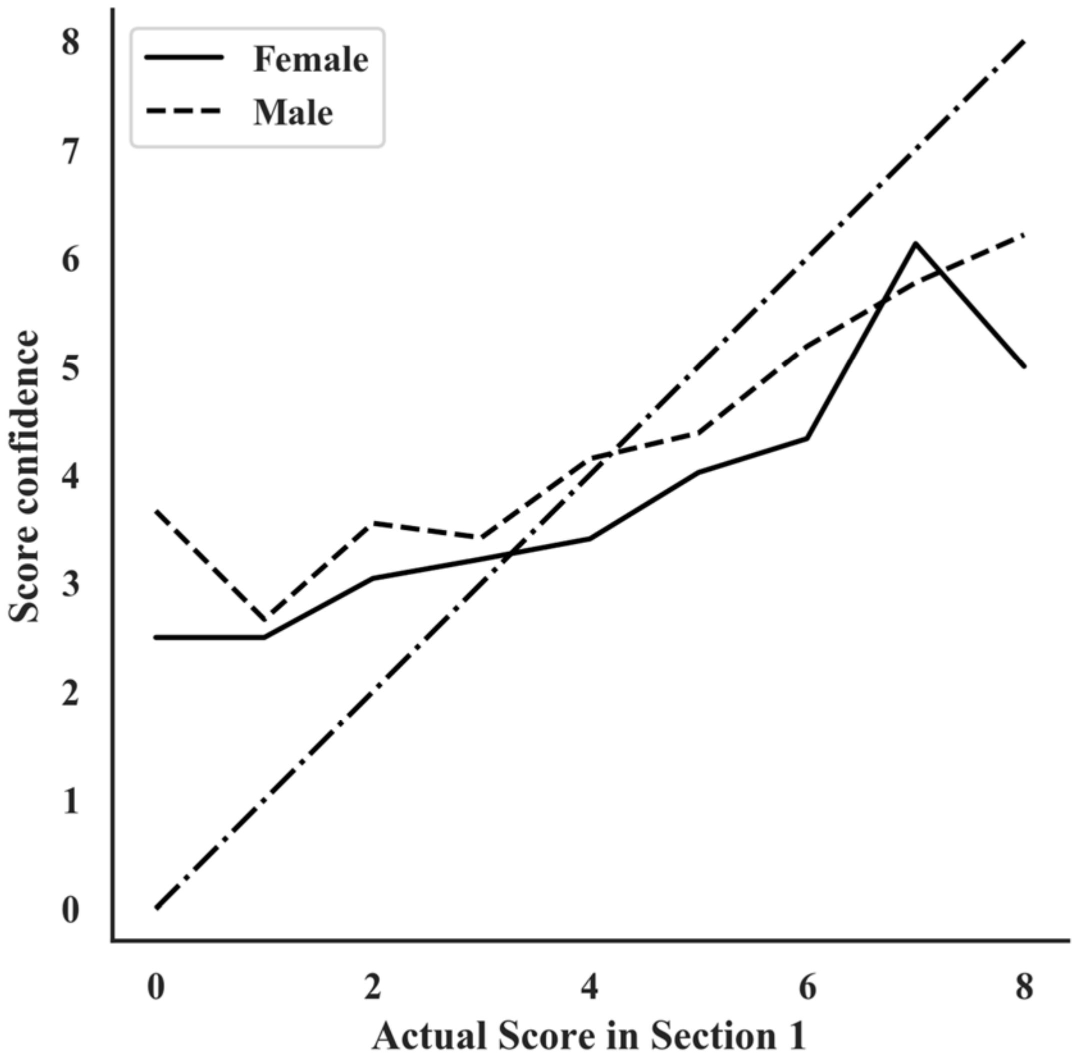


Figure 4: Gender Differences in Tournament Entry with and without Feedback on Average, and by Information Condition about the Gender of Random Match

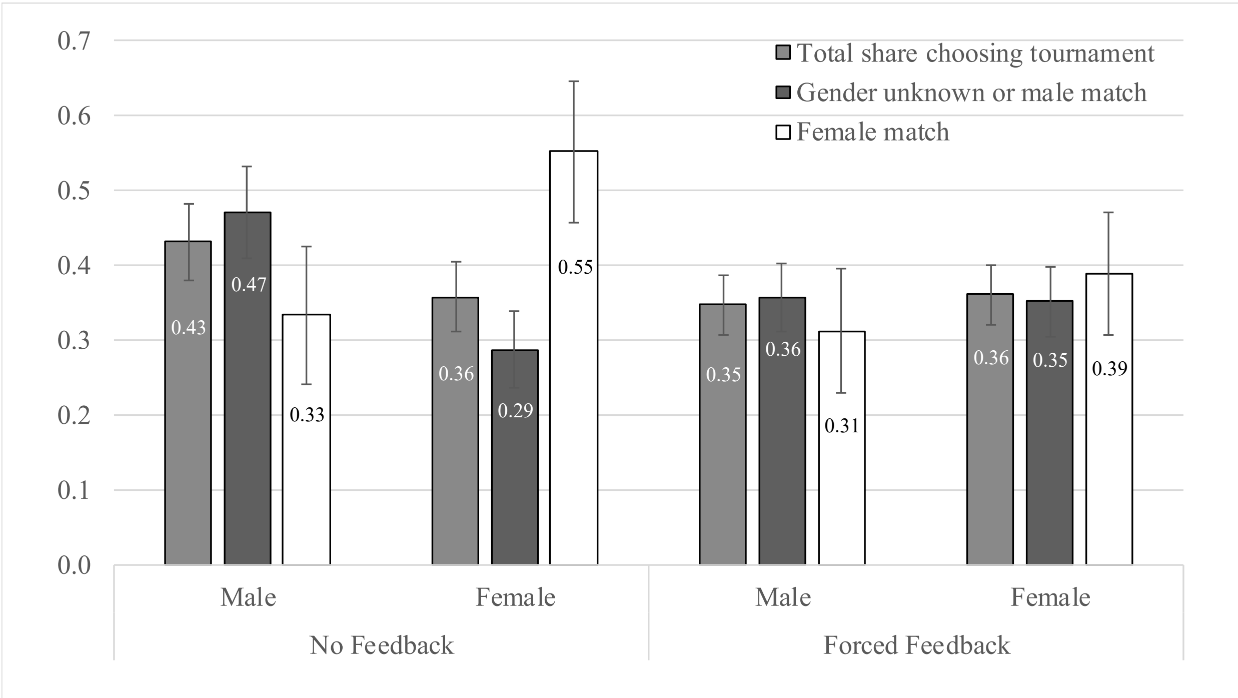
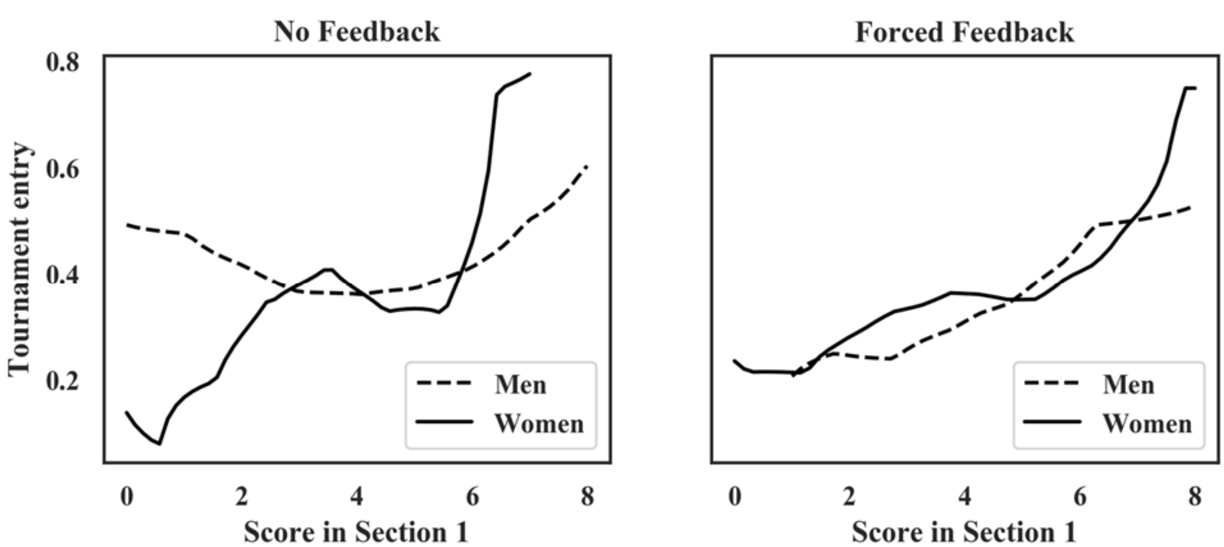


Figure 5: Local Linear Regression of Tournament Entry on Score in Section 1 by Gender and Feedback Treatment Condition



TABLES

Table 1: Treatment Summary

	Gender of Match Unknown (UG)	Gender of Match Known (KG)	Total Subjects
No Feedback (NF)	98	106	204
Forced Feedback (FF)	148	140	288
Optional Feedback (OF)	49	52	101
Total Subjects	295	298	593

Table 2: Summary of Demographic Characteristics and Balance Tests

	Wave 1						Wave 2	Wave 3		
Variables	No Feedback	Forced Feedback	Optional Feedback	NF vs FF p-value	NF vs. OF p-value	FF vs. OF p-value	Forced Feedback	No Feedback	Forced Feedback	NF vs. FF p-value
age	42.23	40.13	40.42	0.14	0.23	0.84	38.2	36.85	36.06	0.64
female	0.59	0.55	0.50	0.59	0.24	0.53	0.5	0.48	0.46	0.77
education										
Less than high school	0.00%	0.00%	0.00%	-	-	-	0.00%	1.00%	0.00%	0.32
High school or GED	13.46%	7.84%	11.88%	0.20	0.78	0.32	21.59%	8.00%	8.16%	0.97
Some College	25.96%	15.69%	21.78%	0.08	0.53	0.25	26.14%	24.00%	22.45%	0.80
2-year college degree	9.62%	14.71%	11.88%	0.25	0.57	0.57	12.50%	7.00%	18.37%	0.02*
4-year college degree	35.58%	46.08%	47.52%	0.11	0.06	0.78	29.55%	47.00%	40.82%	0.38
Master's degree	10.58%	11.76%	3.96%	0.77	0.08	0.04*	7.95%	10.00%	7.14%	0.47
Professional degree	3.85%	2.94%	-	0.73	0.05*	0.08	1.14%	3.00%	1.02%	0.32
Doctoral degreed	0.96%	-	0.99%	0.32	0.97	0.31	1.14%	0.00%	2.04%	0.15
income										
Less than $10,000	3.85%	4.90%	9.90%	0.70	0.08	0.17	4.55%	6.00%	9.18%	0.40
$10,000 - $19,999	11.54%	7.84%	5.94%	0.38	0.17	0.61	6.82%	8.00%	10.20%	0.59
$20,000 - $29,999	15.38%	7.84%	15.84%	0.10	0.88	0.07	18.18%	10.00%	12.24%	0.62
$30,000 - $39,999	6.73%	13.73%	14.85%	0.09	0.05	0.80	15.91%	13.00%	16.33%	0.51
$40,000 - $49,999	16.35%	7.84%	12.87%	0.07	0.52	0.23	11.36%	10.00%	8.16%	0.65
$50,000 - $74,999	22.12%	18.63%	20.79%	0.56	0.88	0.67	21.59%	27.00%	23.47%	0.57
$75,000 - $99,999	13.46%	10.78%	9.90%	0.57	0.46	0.86	9.09%	17.00%	10.20%	0.16
$100,000 - $149,999	6.73%	21.57%	6.93%	0.00*	0.92	0.00*	7.95%	7.00%	8.16%	0.76
$150,000 - $249,999	3.85%	5.88%	0.00%	0.49	0.05*	0.01*	4.55%	2.00%	2.04%	0.98
$250,000 - $499,999	0.00%	0.00%	0.99%	-	0.30	0.31	0.00%	0.00%	0.00%	-
race										
Asian	5.77%	9.80%	1.98%	0.27	0.17	0.02*	6.82%	7.00%	10.20%	0.42
Asian-Pacific Islander	0.96%	1.96%	0.00%	0.54	0.33	0.16	0.00%	0.00%	0.00%	-
Black or African American	4.81%	7.84%	4.95%	0.36	0.94	0.41	3.41%	8.00%	7.14%	0.82
Native American	0.96%	0.00%	0.00%	0.32	0.33	-	0.00%	2.00%	2.04%	0.98
White	84.62%	77.45%	89.11%	0.24	0.17	0.01*	89.77%	83.00%	77.55%	0.34
Other/ Do not wish to disclose	2.88%	1.96%	1.98%	0.67	0.69	0.98	0.00%	0.00%	3.06%	0.08

Notes: Significance levels $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3: Mean Comparisons of Gender Differences in Behavioral Traits

Variable	Male	Female	Diff
Average score in Section 1	4.30	3.79	0.51***
Average bonus in Section 1	0.82	0.71	0.11***
Score Confidence	4.29	3.53	0.76***
Proportion self-evaluating below average	0.41	0.64	-0.23***
Self-Reported Risk Preference	5.85	4.56	1.29***
Number of obs.	239	253	

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.

Table 4: Ordinary Least Squares Estimates of Gender Gaps in Behavioral Traits

Dependent variable	Score Confidence (Self-Reported Score in Section 1)		Self-evaluation of Payment to be Below Average	
	(1)	(2)	(3)	(4)
Female	-0.76***	-0.52***	0.23***	0.18***
	(0.17)	(0.16)	(0.04)	(0.04)
Score in Section 1		0.47***		-0.09***
		(0.04)		(0.01)
Dependent variable mean	3.90	3.94	0.53	0.53
Observations	492	492	492	492
R-squared	0.04	0.23	0.05	0.16

Notes: Robust standard errors in parentheses. All specifications are based on data from no feedback and forced feedback conditions. In Columns 1 and 2, score-confidence is measured on a scale of 0 to 10 (the highest possible score in Section 1 is 8). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Determinants of Tournament Entry Decision by Treatment

Samples	No Feedback		Forced Feedback	
	(1)	(2)	(3)	(4)
Female	-0.27*	-0.25	0.10	0.19
	(0.16)	(0.16)	(0.13)	(0.13)
Male x Score in Section 1	0.01	-0.01	0.07***	0.08***
	(0.03)	(0.03)	(0.02)	(0.02)
Female x Score in Section 1	0.07***	0.06**	0.05**	0.06**
	(0.03)	(0.03)	(0.02)	(0.02)
Risk		0.02		0.06***
		(0.01)		(0.01)
Score Confidence		0.03		-0.02
		(0.02)		(0.02)
Dependent variable mean	0.39	0.39	0.35	0.35
F-test of interactions (p-value)	0.122	0.081	0.611	0.483
Observations	204	204	288	287
R-squared	0.04	0.06	0.05	0.13

Notes: Robust standard errors in parentheses. All specifications are based on data from no feedback and forced feedback conditions. F-test of interactions tests against the null hypothesis that Male x Score in Section 1 is equal to Female X Score in Section 1. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Average Attribution to Own Ability by Gender and Type of Feedback Outcome

Type of Feedback	Male	Female	Difference
Negative reinforcement (Self-Evaluation of Payment Below Average and Payment Actually Below)	74.1	70.8	3.3
Positive surprise (Self-Evaluation of Payment Below Average but Payment Actually Above)	45.3	58.1	-12.8
Negative surprise (Self-Evaluation of Payment Above Average, but Payment Actually Below)	55.4	78.1	-22.76***
Positive reinforcement (Self-Evaluation of Payment above Average and Payment Actually Above)	77.7	64.1	13.65**

Notes: Summary statistics are based on data from waves 1 and 2 of the experiment, forced feedback condition only (wave 3 did not contain the attribution question). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: The Gendered Effect on Attribution of Receiving Negative Feedback Relative to the Original Self-Evaluation Position

Sample	Negative Self-Evaluation		Positive Self-Evaluation	
	(1)	(2)	(3)	(4)
Female	12.83	11.22	-13.65**	-13.42**
	(10.47)	(10.77)	(5.96)	(6.30)
Male x Negative Feedback	28.79***	26.09**	-22.35***	-21.18***
	(9.57)	(10.81)	(5.14)	(7.53)
Female x Negative Feedback	12.65*	10.17	14.06**	15.21*
	(7.05)	(8.21)	(6.55)	(7.92)
Score in Section 1		-0.79		0.5
		(2.31)		(1.77)
Score Confidence		0.43		0.06
		(1.89)		(1.83)
Risk		-1.22		0.17
		(1.29)		(1.01)
Dependent variable mean	67.60	67.60	67.69	67.69
F-test of interactions (p-value)	0.1778	0.1877	<0.0001	<0.0001
Observations	97	96	93	93
R-squared	0.1	0.11	0.21	0.21

Notes: Robust standard errors in parentheses. All specifications are based on data from waves 1 and 2 of the experiment, forced feedback condition only (wave 3 did not contain the attribution question). F-test of interactions tests against the null hypothesis that Male x Negative Feedback is equal to Female x Negative Feedback.
Significance levels: *** p<0.01, ** p<0.05, * p<0.1.