2018

Luck or Skill: How Women and Men Attribute Successes and Failures

Lingjun Xia
lxia@wellesley.edu

Luck or Skill:

How Women and Men Attribute Successes and Failures

Lingjun "Lotus" Xia

Submitted in Partial Fulfillment
of the
Prerequisite for Honors
in Economics
under the advisement of Olga Shurchkov

April 2018

Acknowledgments

First and foremost, I would like to thank Professor Olga Shurchkov, my thesis advisor, and Professor Kartini Shastry for their thoughtful guidance and immense help at every stage of this project. Without them, I would never have been able to accomplish this thesis.

I also want to thank Professor Casey Rothchild and Professor Jeremy Wilmer for showing interests in my project and agreeing to be on my thesis committee. My thanks also to Professor Eric Hilt and the ERS team for their support throughout the year and invaluable comments and advice. I am also grateful to Professors Susan Skeath, Seth Neumuller, and Dan Fetter for deepening my love for and understanding of economics.

I would like to thank the Jerome A. Schiff Fellowship committee for providing me with financial support and recognition for this thesis.

Also, I thank my parents, although they probably do not know I decided to write a thesis. Thank them for supporting me for this great education opportunities and allowing me to freely pursue my interests.

Last but not least, I want to thank Jiafeng Chen for all the emotional support and insightful discussions on this project. This process will be much more stressful without his humorous comments.

Abstract

There is a consensus in economics that a significant gender gap in competitiveness exists, which contributes to substantial gender difference in economic outcomes. Our study uses a controlled online experiment to explore a potential explanatory variable for gender gaps in tournament entry, namely, the gender difference in attribution of feedback. We find that, upon receiving negative feedback, women attribute it to lack of ability, regardless of what self-evaluation they hold initially. On the other hand, men blame bad luck for negative feedback that challenges their positive self-evaluation, and only blame their own ability if their self-evaluation was pessimistic in the first place. We also find that feedback eliminates the gender difference in tournament entry, confirming previous work. The elimination of the gap is mostly attributable to the fact that low-performing men are less likely to enter competition upon receiving feedback. Despite a substantial difference in attribution patterns, we cannot conclude, with our current data, whether feedback attribution is a major explanatory variable for the gender gap in competitiveness.

Table of Contents

Chapter 1. Introduction

Previous research in economics indicates that there is a substantial gender gap in economic outcomes. For instance, Bertrand and Hallock (2001) show that, in a sample of US companies, women account for just 2.5% of top five executive positions. A simple comparison between annual wages of full-time working women and men shows that women earn about 79% of what men earn (Blau and Kahn, 2016).

The potential underlying explanations for the gender gap can be sorted into two broad categories. First, women might be discriminated against directly or differentially treated by employers, superiors, or coworkers, thus having fewer opportunities to achieve better economic outcomes (see Goldin 2014, for an overview). Second, women might self-select into lower-paying jobs and be less likely to undertake lucrative opportunities.[1]

In this paper, we focus on the second explanation. The literature on decision-making has found that differences in behavioral traits partially explain differential sorting into careers and opportunities (see Niederle 2016 and Shurchkov and Eckel 2018 for comprehensive reviews 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 competition when the environment is perceived to favor men, even when there is no significant gender gap in performance. They observe that women would be better off if they competed more.

Two salient explanatory factors of the gender gap in competitiveness are risk aversion and confidence. First, men are more willing to take risks, though the exact gap varies in size and significance (Shurchkov and Eckel 2018). Second, men are more overconfident in their own

[1] It is important to note that these decisions could be made in anticipation of discrimination and may not simply reflect different preferences. In this paper, we focus on the underlying reasons for individual choices, abstracting away from the additional societal pressures brought on by labor market discrimination. .

ability than women (Niederle and Vesterlund 2011). However, the literature has not reached a consensus on whether differential attitudes and reaction toward risk can fully explain the gender gap in tournament entry.[2] Furthermore, the implications of the effect of risk aversion and overconfidence on the gender gap in economic outcomes are unclear: As a society, it may be undesirable to aim to "change" women to become more risk-loving and overconfident. In fact, these traits may lead to other potentially harmful side-effects (see Barber and O'Dean 2001 and Eckel and Fulbruhn 2016, for examples of excess aggressiveness on the part of male participants in financial markets contexts, for example).

This paper aims to elucidate the way in which gendered reactions to feedback may contribute to differences in choices, regardless of risk preferences and confidence. In particular, upon receiving feedback on performance, men and women may differentially attribute the feedback to luck versus ability, update their self-evaluation, and therefore enter competition at different rates.

To study the attribution of feedback, we conducted a controlled online experiment. While observational studies may be preferable from the perspective of external validity, they do not allow the researcher to directly observe attribution of feedback. A controlled online experiment, on the other hand, allows us to exogenously vary the extent to which subjects have access to feedback and measure their attribution of said feedback. Our experimental data complement observational studies of gender differences in preferences and economic outcomes.

In our experiment, participants first work on a real-effort task and learn their payment, determined partially by their own performance and partially by a random "luck" component.

[2] While much of the literature points to overconfidence and inherent competitiveness as primary explanations for the gender gap, with risk preferences playing a somewhat more modest role (Niederle and Vesterlund 2011), recent studies argue that overconfidence and risk preferences can, under some circumstances, fully eliminate the tournament entry gap (Gillen, Snowberg and Yariv 2017; van Veldhuizen 2017).

They are never informed of their score, which represents their real ability. We believe this closely resembles a real-world setting, since in the real-world, people often do not directly observe, but instead must infer, their own relative ability solely based on the outcome. Then, participants are separated into three treatment conditions based upon the feedback they receive: (a) no further feedback; (b) feedback on whether their payment was above or below the group average; (c) option to pay to receive feedback on whether their payment was above or below the group average. Finally, the participants in all three conditions are asked to attribute their payment outcome based on a scale that ranges from pure luck to pure ability (our measure of attribution).

Our first goal is to investigate if there is differential attribution of feedback by gender. Goldin (2013) has noted that women are less likely to major in economics after receiving what they perceive as unsatisfactory grades in introductory classes. Men, on the other hand, are far less likely to be discouraged away from majoring in economics given the same grade. Based on this observation, we hypothesize women to be more likely to attribute negative feedback to a lack of ability, whereas men to be more likely to attribute comparable negative feedback to bad luck. We also explore the patterns of attribution of positive feedback. Intuitively, women may take less credit for positive feedback relative to men. Our results confirm these hypotheses. A summary of our results is presented in the table below.

Upon receiving negative feedback that undermines their positive self-evaluation (Cell A), women attribute this “negative surprise” to their lack of ability, whereas men blame bad luck. This finding is consistent with Goldin’s observation related to major choice. Upon receiving negative feedback that reinforces their negative self-evaluation (Cell B), both men and women attribute the negative outcome to lack of ability. Upon receiving positive feedback that reinforces their positive self-evaluation (Cell C), women do not take credit for the success and are more likely to attribute the positive news to good luck, whereas men's reaction to feedback in this condition is not significantly different from their counterparts who did not receive any feedback. Finally, we do not find consistent gender differences among the subjects who receive positive feedback reinforcing a pessimistic self-evaluation (Cell D).

	Negative Feedback	Positive Feedback
Positive Self-Evaluation	(A) Men: attribute to *luck* Women: attribute to *self*	(C) Men: no obvious reaction Women: attribute to *luck*
Negative Self-Evaluation	(B) Men: attribute to *self* Women: attribute to *self*	(D) No obvious reaction for both men and women

Having established that there is a substantial gender difference in attribution of feedback, our second goal is to test if this gender gap in attribution may explain the gender differences in tournament entry. Intuitively, women who are more likely to attribute failures to lack of ability and successes to good luck may be the ones to fail who properly update their beliefs based on feedback and thus choose not to enter competition, generating a gender gap in tournament entry.

First, we verify that our experimental setup elicits gender differences in tournament entry under the conditions where such disparities are expected to occur. In particular, women are significantly less likely to enter competition in mixed-gender groups in the no feedback condition. This is the case despite certain design differences that result in our tournament being relatively less intensely competitive than competition in NV and other previous studies (see Section 3.4 for reasons behind our specific design choices). Consistent with previous literature, we find that the gender gap in tournament entry is eliminated when the participants receive feedback (see Ertac and Szentes 2010) and when female participants are made aware that they would be facing a woman in competition (see Gneezy, Niederle, and Rustichini 2003 for the effects of single-sex competition on performance in tournaments).

We then proceed to investigate the explanatory power of attribution in driving the gender gap in competitiveness. If attribution is a main explanatory variable bridging the gender gap in tournament entry, then controlling for attribution should eliminate the effect of gender. We find that the differential effect of attribution by gender renders the gender gap in tournament entry insignificant. We also investigate whether gender differences in risk aversion and overconfidence, which are in line with previous literature in our experiment, further explain the gap in tournament entry. The inclusion of these behavioral traits in the analysis indeed further reduces the gender gap in tournament entry, making it both statistically and economically insignificant.

In addition, provided that feedback eliminates the gender gap in tournament entry, we explore how men and women self-select into tournament in the absence and in the presence of feedback. When there is no feedback on one's relative performance, higher-performing women are more likely to select into competition, whereas men are equally likely to enter regardless of performance. This positive selection into tournament provides a tournament-entry counterpart of the findings by Exley, Niederle, and Vesterlund (2016), who observe positive self-selection into *negotiation* by women, but not by men. Upon receiving feedback, however, men and women exhibit similar patterns of tournament entry – higher score predicts a higher probability of entry. This implies that feedback eliminates the tournament entry gap by correcting previously suboptimal selection choices of low-performing men.

Even though our experimental design is stylized and simplified when compared to real-world situations, our findings have important implications. Although willingness to enter competition is sometimes associated with better economic outcomes, it is not always optimal to force women to compete. Women in our experiment correctly sort into tournament based on their score even in the absence of feedback, whereas men tend to blindly enter competition. In general, timely feedback on performance helps to eliminate the gender difference in competitiveness, but the type of feedback is important and may lead to different consequences. The best feedback for women should emphasize the role of ability.

Finally, because the tournament in the second stage of our experiment is identical to that in the first stage, our design abstracts away from additional positive aspects of persisting in competitive environments, such as human capital accumulation associated with progressing in a job or picking a particular competitive major of study. In those more realistic settings, misattribution might play an additional role that we cannot observe in our setting. In financial markets, unlucky high-ability female traders may misattribute losses to lack of ability rather than to bad luck and would thus be deterred from entering competitive environments. Closing the gender gap in attribution could therefore help to eliminate the gap in investment behavior.

The rest of the paper is organized as follows. Chapter 2 presents a brief literature review on gender difference in competitiveness and reaction to feedback. Chapter 3 describes the experiment design. Chapter 4 presents the results of the experiment. Chapter 5 concludes with some implication of the results and discusses potential future research.

Chapter 2. Literature Review

It has been widely documented that there is a gender gap in economic outcomes. A simple comparison between annual wages of full-time working women and men shows that women earn about 79% of what men do (Blau and Kahn, 2016). This gap in outcomes is accounted for by two channels: First, women tend to shy away from riskier but more lucrative fields (Kleinjans, 2008). Second, women achieve a lesser status relative to men within the same occupation. They may begin with similar entry-level jobs after graduation, but as their seniority increases, men are more likely to obtain executive positions and earn higher wages then women (Bertrand, Goldin, and Katz 2010, Crespo, Simoes, and Moreira 2014).

Two factors might explain the existence of a gender gap in occupation choices and in status achieved. First, women might simply be differentially treated by employers, superiors, or coworkers, contributing to the gender gap in within-occupation status (Wolfers 2015, Sarsons 2017). Second, women might be less likely to search for or to undertake opportunities that might lead to career advancement or economic improvement. Women usually choose to enter less financially lucrative or socially prestigious professions (Buser, Niederle, and Oosterbeek 2014, Goldin 2017). Women are also less likely to invest in financial market, which is another opportunity that is conducive to economic outcomes: women display lower levels of financial literacy than men (for example, Lusardi and Mitchell, 2008; Almenberg and Dreber, 2012; de Bassa Scheresberg, 2013; Agnew and Harrison, 2016, Bucher-Koenen et al. 2016), manifest less financial market participation (Almenberg and Dreber 2012, Halko, Maustia and Alanko 2012), and exhibit less retirement planning (Lusardi and Mitchell, 2008). However, the gap in willingness to undertake opportunities does not necessarily reflect a gap in competency.

Although women participate less in the stock market, those who do participate do better than men (Barber and Odean 2001).

A caveat is that the two factors may reinforce each other and might be empirically confounding. If differential treatment exists and lowers the expected utility of a career move for women, then we might expect women to be less likely to pursue the opportunity. The existence of this interacting channel colors empirical analysis, as we may not be able to tell apart which factor is in action.

This paper mainly focuses on the second factor, namely the gender gap in *decision making*, where women may be less likely to pursue financially lucrative opportunities.

2.1 Competitiveness

There is a consensus that the gender gap in decision-making is broadly attributable to a gap in competitiveness, as many opportunities are competitive in nature. Researchers have found that women and men differ in their reaction towards competition: men are more willing to enter competitions in the first place (NV, Shurchkov and Eckel 2018). Within competition, men tend to outperform women, while this difference in performance disappears in non-competitive environments (Croson & Gneezy 2004, Shurchkov and Eckel 2018).

There is a significant gender difference in the decision to enter competition. In their seminal paper on this topic, NV used a stereotypically male task that involved adding up sets of five two-digit numbers within a certain time limit. They then compared the probability of men and women to self-select into a winner-take-all tournament payment scheme, where each participant's score is compared with three other participants', and only the winner in the group gets an extra bonus. They found that men select the winner-take-all payment scheme twice as often as women. This gap is not explained by participants' performance – some well-performing women also selected piece-rate payment and, as a result, hurt themselves financially.

It is important to note, however, that the gender stereotype associated with the task may elicits different levels of gender gap. While women are found to be less competitive in stereotypically male tasks (for example, Dargnies, 2012; Brandts, Groenert, and Rott, 2014; Buser, Niederle and Oosterbeek, 2014), some studies noted that stereotypically female tasks rarely induce gender difference in tournament entry (Kamas and Preston 2009; Shurchkov 2012).

In terms of gender gap in performance *within* competition, the literature has less consensus. Some studies using the addition task in NV's design did find such a gender gap in performance within tournament (for example, NV, Healy and Pate 2011; Buser, Niederle, and Oosterbeek 2014). However, studies examining other stereotypically male-favoring tasks did find a difference in performance under tournament-based payment scheme (for example, Gneezy, Niederle, and Rustichini 2003, Gneezy and Rustichini 2004, Shurchkov 2012, Schram, Brandys, and Bërxhani 2016).

So far, we have established that there is a gender gap in attitudes towards and reactions to competition. The natural question now is: What explains the gap in competitiveness? Two salient factors are *risk aversion* and *confidence*.

The consensus in the literature is that women exhibit greater risk aversion (Eckel and Grossman 2002, Charness and Gneezy 2011) and lower levels of confidence (Beyer 1990, Barber and Odean 2001). Some studies found that risk aversion explains the gender gap (Balafoutas, Kerschbamer, and Sutter 2011, Gillen, Snowberg, Yariv (2017)) while others argued that the gender gap persists even after controlling for risk preferences (NV). In terms of confidence, NV argued that overconfidence of men explains a portion of gender difference in tournament entry. However, a significant gap remains after controlling for confidence level.

We aim to elucidate a third explanatory variable for the gender gap in competitiveness: attribution of feedback. In the real-world, the outcome one obtains is usually determined by one's own ability and an external luck component. Upon receiving feedback on the outcome, men and women may attribute the feedback differently to ability and luck, and therefore have a different tournament entry pattern.

2.2 Feedback

There has been an extensive literature on the reaction to feedback in psychology. The main finding is *self-attribution bias*: a tendency for people to attribute success to their own ability but ascribe failure to some external factors to maintain self-esteem (Miller & Ross, 1975; Mezulis et al. 2004; Aronson et al. 2013). Some economic studies have replicated this self-serving interpretation of feedback. People prefer feedback that suggests an external cause of poor performance to feedback that suggests poor ability (Liden and Mitchel, 1985); people prefer to attribute good performance to their own ability (Hoffman and Post 2014). Positive feedback to performance is also interpreted as more informative and elicits a stronger behavioral reaction than negative feedback (Mobius et al. 2014, Eil and Rao 2011, Love, Love, and Northcreaft 2010). In addition, people are more likely to agree with and to react to the feedback that is consistent with their self-appraisal (Korsgaard 1996).

There is also gender-specific reaction to feedback. Mobius et al. (2014) found that, provided with a feedback, positive or negative, women revised their beliefs about their performance less than men do. Berlin and Dargnies (2016), however, found the opposite results:

women reacted more strongly than men to both positive and negative feedback – they became more confident after receiving positive feedback and less confident after negative feedback. Note a major difference between the two studies that might explain the opposite results: The feedback in Mobius et al. (2014)'s study has a 75% accuracy, but the feedback in Dargnies (2016)'s study is guaranteed to be correct. It is not hard to imagine that people react more strongly to accurate feedback than to not fully accurate ones. Another study by McCarty (1986) notes that women are less confident than men, regardless of what feedback they receive.

Two studies linking reaction to feedback to tournament entry are worth more attention for purpose of this study. Ertac and Szentes (2010) found that performance feedback reduced the gender gap in competitiveness. They closely followed NV's design and observed that, when there was no performance feedback, women were less likely than men to choose a winner-take-all payment scheme. However, upon receiving a feedback about the highest score in the group, there was no gender difference in decision to compete. Another study by Brandt, Groenert, and Rott (2014) studied the effect of *advice* on entry of competition. Instead of an objective performance feedback, participants in their study received a piece of advice from a more informed advisor regarding what payment scheme they should choose for the next round of problem solving. Note that a normative advice is potentially different from a descriptive feedback. One can follow advice from other people without being able to come up with the same conclusion from the feedback. The results show that more high-performing women self-selected into tournament payment after being advised to do so, but the gender gap in tournament entry remained.

Our study is innovative and contributes to the literature in two ways: 1. Most existing studies focus on feedback of performance (ability). While these studies provide a good starting point, they are not realistic in representing the real-world situation, since feedback in real-world is usually determined by both ability and luck. In our experiment, the feedback participants receive depends on both their performance in a task and a random luck component. 2. Benefitting from our design of feedback, we can examine the *channel* through which feedback influences competitiveness, namely, men and women's differential attribution of feedback to luck and ability. To the best of our knowledge, we are the first to study the mechanism between feedback and competitiveness.

Chapter 3: Experimental Design

The goal of this study is to investigate whether women and men differ in how they attribute positive feedback (successes) and negative feedback (failures) to their own ability versus to luck, and, if the gender difference in attribution exists, to examine whether it explains the gender gap in tournament entry decisions.

We use a controlled online survey experiment in order to measure attribution and control the level of feedback participants receive. Participants complete a section of problem solving, get feedback, and attribute their successes and failures. They then choose between a piece-rate payment scheme and a tournament-based payment scheme, and proceed to complete another section of problem solving. *Attribution* is measured by participants' assignment of the positive or negative feedback to their own ability versus luck, scaled from 0 to 100. A score of 100 means attributing the feedback entirely to ability, and a score of 0 means attributing entirely to luck. *Tournament entry* is elicited through subjects' choice of tournament-based payment scheme for the second round of problem solving. We also gauge participants' *confidence* level by two measures: We ask participants to estimate their score in the first problem solving section (henceforth *score-confidence*) and to evaluate whether their payment from the first section is above or below average (henceforth *payment-confidence*). *Risk* preference is measured as a self-reported risk level by subjects, scale from 1 to 10, where 10 indicates an extreme willingness to take risks, and 0 means an extreme risk aversion.

Our design differs from the existing literature in the following two ways. First, most experiments that focus on behavioral traits use a traditional laboratory setting, while we explore an emerging online participant pool via Amazon Mechanical Turk (MTurk). Research has shown that workers on MTurk exhibit similar heuristics and biases as subjects from tradition sources.

They also pay attention to the instructions as least as much as traditional subjects (Paolacci, Chandler, and Ipeirotis, 2010; Germine et al. 2012). The only concern is that workers on MTurk are usually more experienced than subjects in traditional labs, since they may have constantly participated in experiments.

Second, unlike NV and other subsequent studies (e.g., Shurchkov 2012, Kamas and Preston 2012, Brandts, Groenert, and Rott 2014) that use a winner-take-all payment scheme with groups of four competitors, we adopt a less competitive tournament payment scheme, where participants compete against just one other competitor and where the loser is still guaranteed to receive a certain payment. The winner-take-all design of previous literature is realistic when modeling some scenarios, such as those where only one among many gets a promotion opportunity; however, when taking a broader view, our more continuous payment scheme is more realistic, since we may expect that, *on average*, the expected reward is continuous in, say, human capital.

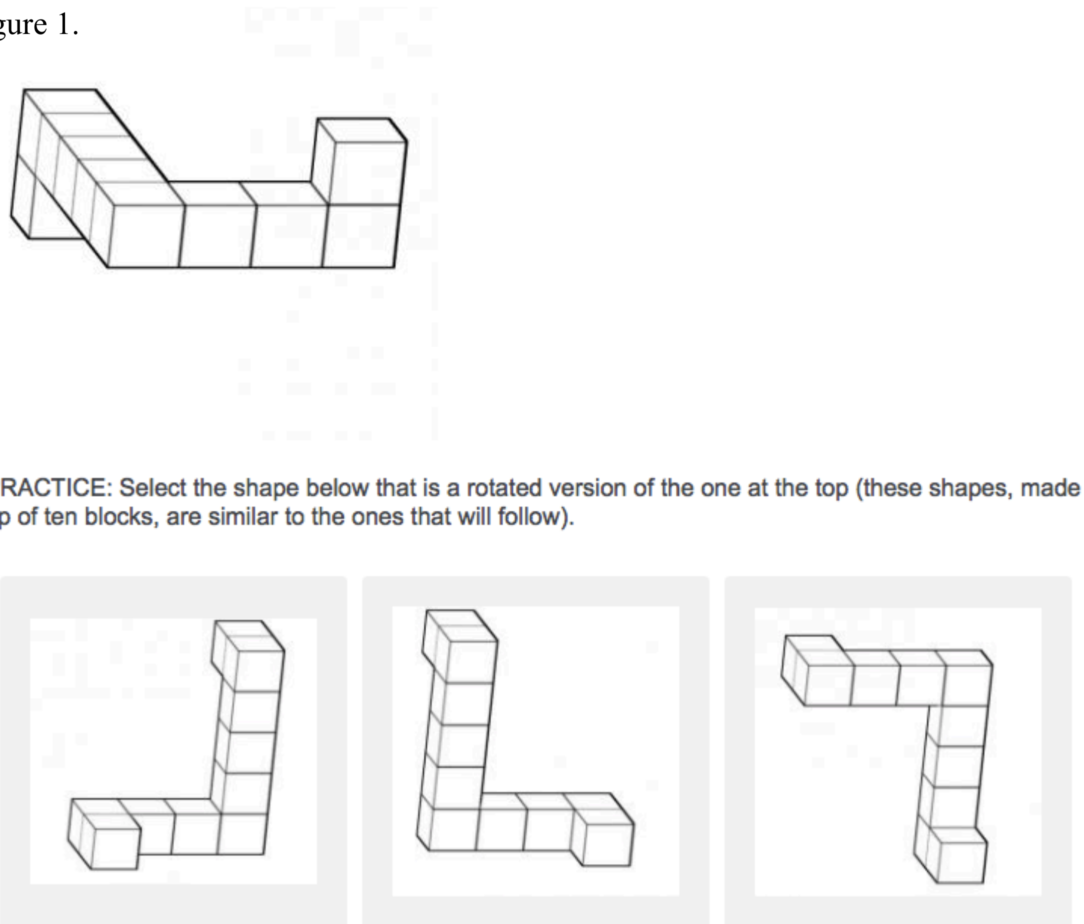
3.1. The Task

The experiment involves two sections of problem solving. For both sections, participants work on Mental Rotation Test (MRT).[3] It is a test in which participants see a target three-dimensional shape made of 10 cubes and identify the rotated version of the target shape among three choices (Figure 1 shows a sample problem).

---

[3] The MRT questions we use in our experiment is 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 take out one of the correct choices for each target shape, and remove the penalty. Therefore, in our MRT, participants choose one out of three choices. They get one point for each correct answer and zero point for each wrong or blank answer.

Figure 1.



This type of task is selected based on two criteria: (1) the task is stereotypically perceived to favor men, and (2) there is no actual difference in performance between men and women.[4] Our pilot studies show that, when asked to evaluate if women or men would score higher on the MRT, significantly more participants selected men ($p < 0.05$). Pilot results also show that the actual scores do not differ significantly between men and women ($p = 0.13$). Note that we depart from the literature that typically uses an addition task, because it is impossible to prevent online subjects from using a calculator. Similarly, in the pilot, we also experimented with a

[4] Iriberri and Rey-Biel (2012) found that men and women do not perform significantly differently in MRT when they did not know the gender of their rivals. However, the MRT questions adopted in their experiment are different from ours. They showed participants two shapes and asked them to indicate if the two shapes are rotated versions or mirror images of each other, whereas we ask participants to choose, among three choices, a rotated version of a target shape. We did not find a significant gender gap in performance in pilot study. However, in the main study, men significantly outperform women. The psychology literature (e.g. Feng, Spence, and Pratt 2007) has also shown that MRT consistently elicits gender difference in performance.

stereotypically female-favoring task, namely, the anagram task. However, we found that participants were able to, and in fact did, use online search engines to achieve perfect scores.

Before completing the first section of problem-solving, each participant is given the MRT instructions, as well as a practice problem. They cannot advance to the real problem-solving section until they correctly solve the practice question.

Within each problem-solving section, 8 MRT questions are presented on one web page. The order of the questions is randomized for each participant. Participants have exactly 2.5 minutes to complete as many questions as possible.

After completing the task, participants estimate their scores in this section of problem-solving (score-confidence); we do not explicitly tell participants the total number of questions in this section.

3.2 The Treatments

We adopt a 2x3 design where the two treatments are the gender information of the opponent in Section 1 and the feedback information the participant receives.

After participants finish the first section of MRT, they are informed that their score is compared with a randomly selected participant who had previously taken the same test, and their payment is calculated in the following way: If their score is higher than that of their random match, then they get 20 cents for each correct answer; if their score is lower than that of their random match, then they get 15 cents for each correct answer.

Afterwards, we tell them the payment they earned from the first section of MRT. The exact wording of the payment information varies depending on which gender treatment the participants are assigned to.

3.2.1 Match with unknown gender

In this treatment group, participants see their payment without learning anything about the gender of their random match:

> "*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 use a gender noncommittal pronoun "his/hers" to express that the random match might be either gender.

Immediately after participants learn their payment, they are asked to evaluate whether their payment is above average or below average (i.e. payment-confidence). Their evaluation in this treatment establishes a baseline level of payment-confidence.

3.2.2 Match with known gender

In this treatment group, we reveal the gender of the randomly selected match when informing participants of their payment from the first problem-solving section.

"*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 randomly match.[5] The gender specification is made implicit but repeated twice ("female participant", and "her", for example) to prevent participant from guessing the intention and to ensure they pick up the cue.

Then we elicit participants' payment-confidence, later compared with the baseline payment-confidence in the unknown-gender condition. This treatment is intended to examine

[5] One participant from the pilot study indicated that neither "she" or "him" are preferred pronoun. We dropped this participant as a potential randomly.

whether people react differently when they know the gender of their match. We speculate that women might be less payment-confident when they realize that they are competing with men. On the other hand, men may be more payment-confident when they see that their opponent is a woman.

Afterwards, participants are further assigned to three feedback conditions.

3.2.3 No feedback condition

In this condition, we do not give any feedback on participants' payment or score in the first section of MRT. Participants are asked to attribute their payment to luck or ability based on their payment-confidence – for example, if the participant is payment-confident, we ask to what extent she would attribute her purportedly *above average* payment to ability versus luck; if the participant is *not* payment-confident, she would attribute her purportedly *below average* payment.

3.2.4 Forced feedback condition

In this condition, after participants report their payment-confidence, we give them feedback on whether their payment is *actually* above or below average. Participants are then asked to attribute their below or above average payment (based on the feedback) to ability versus luck.

3.2.5 Optional feedback condition

In this condition, participants have an opportunity to give up some of their payment in exchange for feedback on whether their payment actually exceeds the group average.

We first present a few "prices" of the feedback. Participants then indicate the maximum price they are willing to pay for the feedback about their relative payment. Afterwards, we randomly draw a price from the presented price options. If the participant indicates that she is willing to pay the randomly drawn price, then the price is subtracted from her payment, and she receives feedback about whether her payment is actually above or below average. If the participant indicates that she is not willing to pay, then she does not get any feedback.

For participants who do not get any feedback, their attribution is with respect to *their own payment-confidence*. For participants who get feedback, their attribution is with respect to their *actual* relative performance.

We use a 2x3 treatment design. The following table shows the total number of participants in each treatment condition for each wave.[6]

	Gender of Opponent Unknown	Gender of Opponent Known
No feedback condition	Wave 1: 46 Wave 3: 52	Wave 1: 58 Wave 3: 48
Forced feedback condition	Wave 1: 58 Wave 2: 44 Wave 3: 46	Wave 1: 44 Wave 2: 44 Wave 3: 52
Choice of feedback condition	Wave 1: 49	Wave 1: 52

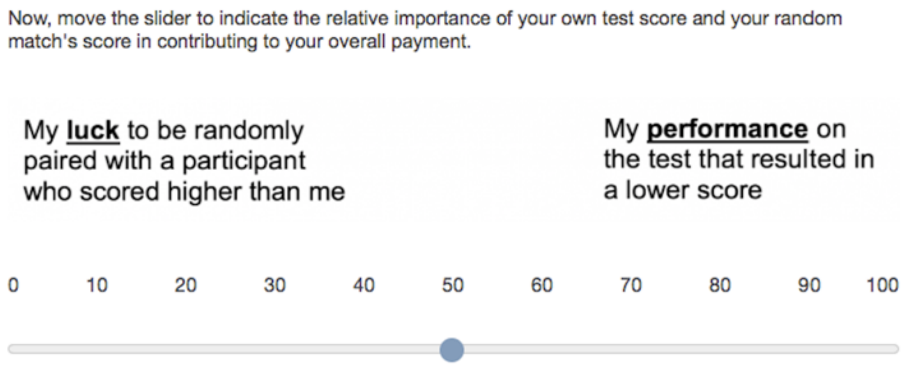
3.3. Attribution of success and failures

After receiving these feedback conditions, participants move a slider to assign relative importance to the two components that contribute to their payment, luck and ability (see Figure 2).

For participants who attribute a negative outcome (whether the negative outcome is perceived or actual), the two components are (1) "my performance on the test that resulted in a *lower* score," and (2) "my luck to be randomly paired with a participant who scored *higher* than me." For participants who attribute a positive outcome, the two components are (1) "my performance on the test that resulted in a *higher* score," and (2) "my luck to be randomly paired with a participant who scored *lower* than me."

[6] We conducted three waves of the experiment in total. See Section 3.5.

Figure 2.



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.

3.4. Tournament entry

Before the second section of problem solving began, participants choose between a tournament-based payment scheme and a piece-rate payment scheme:

- Piece rate: Participant gets 17.5 cents for each correct answer, regardless of anyone else's score.
- Tournament payment: We randomly pair the participant with a match from the other group. If the participant's score is higher than the match's, then the participant gets 25 cents for each correct answer. If the participant's score is lower than the match's, then the participant gets 10 cents for each correct answer.

Participants' bonus payment in the second round of problem solving is calculated based on the payment scheme they choose.

Note that we did not choose a winner-take-all payment scheme: This is different from the design of many previous literatures, including NV. Granted that the winner-take-all design is realistic when modeling some scenarios, such as promotion, we believe our design is more realistic when taking a broader view of the real world. For example, Alice and Bob compete for gold and silver medals, and Alice gets the gold medal and Bob gets the silver medal. If we only look at who gets the gold medal, then reward is indeed winner-take-all, but the bird's eye view suggests that our payment scheme is more reasonable in this setting.

3.5. The procedure

The experiment was conducted online via MTurk. Participants were directed to our experiment pages if they chose to participate. A consent form was presented to the participants prior to the beginning of the experiment. Participants were encouraged to print out and save a copy of the consent form. Participants who indicated they are willing to participate then advanced to the general instruction of the experiment.

Participants were informed in the beginning of the experiment that there would be an attention checking question in the survey. Participants who failed to answer this question correctly were directed to the ending page. They did not get any payment for participating in the experiment.

For each problem-solving section, time ran out automatically. Participants could not advance to the next page before the end of 2.5 minutes. Once the time ran out, the webpage automatically jumped to the next page.

At the end of the experiment, each participant filled out a short questionnaire. The questionnaire asked participants questions on demographics, beliefs, and preferences. In particular, participants reported their age and gender. To gauge the gender perception associated with MRT, we also asked participants if they think women or men are more likely to get a high score in the problem-solving section. We also elicited participants' risk preference by asking them to self-report, on a scale of 1 to 10, how willing and prepared they are to take risks. Research has shown that the self-reported risk level correlates significantly with experimental outcome of risk. It is also a common practice to elicit risk preference through a hypothetical question (Kagel and Roth, 2016).

After participants finished the experiment, they were paid a base payment of $0.5 dollars for completion. The final payment, including their bonus payment earned in 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 experiment was about 11 minutes.

We conducted three waves of data collection. In the first wave, we administered the full experiment design as described above. MTurk recorded 308 completed responses. 1 participant failed to pass our attention check. In the second wave, we corrected a mistake made in the first round, and included only the forced feedback condition, as the error affected the forced feedback condition the most.[7] 88 valid answers were recorded. In the third wave, we excluded the optional feedback condition, because very few people actually opted into receiving feedback, and it is difficult to compare those who sorted into receiving feedback against those who did not. We also dropped the attribution question, concerned that attribution may prime tournament entry. A total of 198 valid answers were collected in this round.

[7] A coding error in Wave 1 of 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 are able to use the data in our analysis.

3.6 Variables

The table below presents a list of key concepts and variables with their respective definitions:

Variable or concept	Description
Score-confidence	Participants' self-evaluation of the number of questions they answer correctly in Section 1. The value ranges from 0 to 10.
Payment-confidence	A participant is *payment-confident* if she rates her payment as above average. She is not *payment-confident* otherwise.
Score in Section 1	The actual score a participant receives in Section 1, ranging from 0 to 8.
Attribution	A value from 0 to 100 where 0 represents that the participant attributes her outcome entirely to luck and 100 represents that she attributes entirely to performance.
Gender of match	The information that the participant receives regarding the gender of her randomly selected opponent. The gender of her match is one of *male*, *female*, or *gender-unknown*.
Feedback positivity	A feedback is *positive* if it indicates that payment is above average, and *negative* otherwise.
Feedback condition	There are three feedback conditions: *no feedback*, *forced feedback*, and *optional feedback*.
Tournament entry	An indicator variable for whether a participant chooses the tournament-based payment scheme.
Risk	A self-reported value of risk preference ranging from 0 to 10, with higher values indicating more willing and more prepared to take risks.

Chapter 4. Analysis of Experimental Results

4.1 Summary statistics

Table 1 reports the summary statistics of demographic information and performs balance tests between treatment groups based on demographics. Only 9 out of 94 comparisons are significant at the 0.05 level. Provided that all the procedures are randomized in the survey, we proceed to analyze the results.[8] 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 towards white and from African-Americans and other minorities (US Census). The income makeup in our study is similar to the national distribution (US Census).

Table 2 reports summary statistics of key experimental variables by treatment and gender. In the main study, men outperform women in Section 1. In this section, there is no pressure of competition, since participants do not learn the payment scheme until after they completed Section 1. Therefore, their performance in this section should replicate results in ordinary MRT studies. Indeed, there is a consensus in psychology that MRT consistently elicits gender difference in performance (Masters and Sanders, 1993). Our results in the main experiment are different from the results in the pilot study, which found no significant difference between men and women in the MRT. A potential cause of the difference is the relatively small sample of participants in the pilot study.

---

[8] In Appendix 5, we show that our results are robust to controlling for demographics that are not balanced across treatments.

Table 1. Summary of Demographics Variables

	Wave 1						Wave 2	Wave 3		
Variables	No Feedback	Forced Feedback	Optional Feedback	No Feedback vs. Forced Feedback (p-value)	No Feedback vs. optional Feedback (p-value)	Forced Feedback vs. optional Feedback (p-value)	Forced Feedback	No Feedback	Forced Feedback	No Feedback vs. Forced Feedback (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: * $p < 0.05$

Table 2: Summary of Key Experimental Variables

	Wave 1									Wave 2			Wave 3					
	No Feedback			Forced Feedback			Optional Feedback			Forced Feedback			No Feedback			Forced Feedback		
Variables	Male	Female	Diff	Male	Female	Diff	Male	Female	Diff	Male	Female	Diff	Male	Female	Diff	Male	Female	Diff
Average score in Section 1	4.23	3.57	0.66*	4.48	3.93	0.55	4.38	3.47	0.91***	4.56	4.28	0.28	4.15	3.56	0.59	4.13	3.69	0.44
Average score in Section 2	4.37	3.85	0.52	5.07	4.09	0.98**	4.82	3.57	1.25***	5.13	4.37	0.76*	4.46	3.85	0.61	4.62	4.40	0.22
tournament-based payment	4.17	4.95	-0.78	5.24	4.47	0.77	5.06	4.33	0.73	6.18	4.95	1.23**	3.91	3.80	0.11	4.65	4.56	0.09
piece-rate payment	4.52	3.36	1.16**	4.90	3.95	0.95*	4.70	3.33	1.37***	4.50	3.92	0.58	4.90	3.89	1.00*	4.61	4.30	0.31
% entering tournament	0.42	0.31	0.11	0.35	0.27	0.08	0.32	0.24	0.08	0.38	0.44	-0.06	0.44	0.42	0.03	0.32	0.40	-0.08
Average bonus in Section 1	0.81	0.66	0.15*	0.87	0.74	0.13*	0.84	0.64	0.20***	0.87	0.83	0.05	0.79	0.66	0.13	0.80	0.69	0.10
gender-unknown match	0.81	0.63	0.18	0.87	0.68	0.18*	0.82	0.67	0.15	0.83	0.86	-0.02	0.82	0.62	0.20*	0.86	0.63	0.24**
Female match	0.86	0.79	0.07	0.72	0.95	-0.23	0.66	0.56	0.1	0.78	0.75	0.03	0.77	0.64	0.13	0.68	0.72	-0.04
Male match	0.74	0.56	0.18	0.97	0.67	0.30*	1.02	0.68	0.34*	1.04	0.87	0.17	0.72	0.74	-0.02	0.82	0.81	0.01
Average bonus in Section 2	0.82	0.78	0.05	1.00	0.77	0.23**	0.96	0.68	0.28***	1.04	0.89	0.15	0.85	0.75	0.10	0.89	0.85	0.04
Average bid for feedback	-	-	-	-	-	-	0.06	0.07	0.01	-	-	-	-	-	-	-	-	-
Score-confidence	4.16	3.56	0.61	4.57	3.30	1.26***	4.48	3.78	0.70**	4.36	3.60	0.75*	4.44	3.42	1.02***	3.94	3.82	0.12
Average attribution	63.35	60.05	3.3	64.96	71.38	-6.42	67.72	59.16	8.56*	68.20	65.07	3.13	-	-	-	-	-	-
% participants not payment confident	0.53	0.75	-0.22**	0.35	0.63	-0.28***	0.42	0.65	-0.23**	0.38	0.67	-0.30***	0.35	0.52	-0.17*	0.47	0.60	-0.13
Risk	5.67	4.31	1.36***	5.62	4.29	1.34***	5.35	4.31	1.04**	6.16	4.49	1.67***	6.10	4.65	1.45***	5.68	5.22	0.46

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

For the rest of this paper, we pooled the three waves when conducting analyses. Separate analysis based on each wave is shown in Appendices 2 through 4.

4.2 Gender gaps in risk aversion and confidence

We first verify that our experimental design elicits gender gaps in behavioral traits such as risk aversion or confidence. Table 3 presents the results of a simple OLS regressions comparing men and women along these traits.

Table 3. Gender Gap in Risk Preference, Score-confidence, and Payment-confidence

	(1)	(2)	(3)	(4)	(5)
dependent variables	Risk	Score-confidence	Score-confidence	Proportion not payment-confident	Proportion not payment-confident
female	-1.25***	-0.75***	-0.49***	0.23***	0.17***
	(0.19)	(0.15)	(0.14)	(0.04)	(0.04)
score in Section 1			0.44***		-0.09***
			(0.04)		(0.01)
Dependent variable mean	5.12	3.94	3.94	0.53	0.53
Observations	590	593	593	593	593
R-squared	0.07	0.04	0.21	0.05	0.16

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
All specifications are based on data from all three waves. In Column 1, risk is on a scale from 1 to 10, with 0 representing extreme unwillingness to take risks and 10 representing extreme willingness to take risks. In Column 2 and 3, score-confidence is measured on a scale of 0 to 10. As a comparison, the highest possible score in Section 1 is 8.

In terms of risk preference, men self-report to be more willing and more prepared to take risks than women (Column 1 of Table 3). Provided that self-reported risk preference highly correlates with actual risk aversion uniformly across gender, our finding is consistent with the literature (see for example Charness and Gneezy 2011).

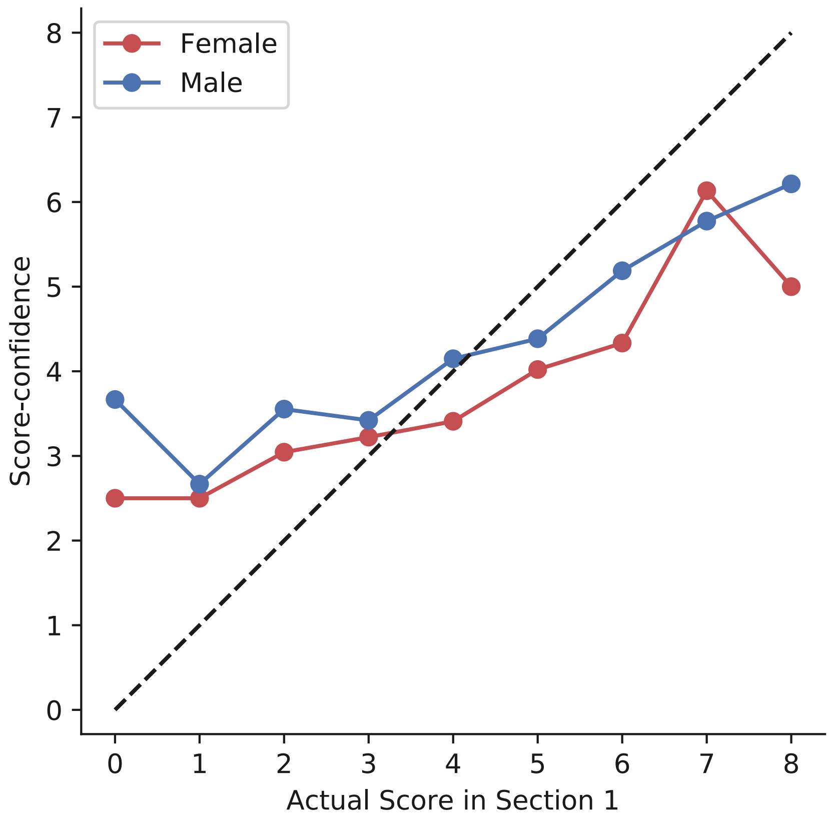
In terms of confidence, our results also confirm previous findings (for example, Beyer, 1990) with both measures of confidence showing that men self-evaluate more highly than women.

First, men are more score-confident than women (Column 2 of Table 3). Out of 8 questions, men on average estimate themselves to solve 4.32 questions, whereas women estimate 3.57 ($p < 0.01$). The difference remains significant after we control for the actual score in Section 1 (Column 3 of Table 3). Second, men are more payment-confident than women: 64% of women are not payment-confident, whereas only 42% of men hold the same belief (Column 4 of Table 3). The difference remains significant after we control for the score in Section 1 (Column 5 of Table 3).

We then examine the two measures of confidence, score-confidence and payment-confidence, separately in more detail. Figure 3 plots the relationship between participants' expected and actual scores in Section 1. 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 (for women: Mean(actual score) = 5.74; Mean(expected score) = 4.48; $p < 0.01$. For men: Mean(actual score) = 6.05; Mean(expected score) = 5.11; $p < 0.01$). Participants of both genders with lower-than-median performance tend to overestimate their score (for women: Mean(actual score) = 2.23;

Mean(expected score) = 3.02; $p < 0.01$. For men: Mean(actual score) = 2.35; Mean(expected score) = 3.44; $p < 0.01$).

Figure 3. The relation between estimated and actual scores



Notes: This graph is the sample average score-confidence, conditional on actual score in Section 1, separated by gender.

In addition, we use a linear probability model below to test how one's own gender and the gender of the opponent affect participants' payment-confidence.[9]

$$\begin{aligned} Not\ payment\ confident \\ &= \beta_0 + \beta_1\ female + \beta_2 genderUnknown\ match + \beta_3 male\ natch \\ &+ \beta_4 femXgenderUnknown\ match + \beta_5 femXmale\ match + \epsilon, \end{aligned}$$

We show the results in Table 4.

[9] Results are robust to a logistic specification and estimates are available upon request.

Table 4. Influence of Gender of the Opponent on Payment-confidence
Dependent variable: not payment confident

	(1)	(2)
female	0.09	0.08
	(0.08)	(0.08)
gender-unknown match	-0.11	-0.08
	(0.07)	(0.07)
male match	-0.11	-0.05
	(0.08)	(0.08)
femaleXgender-unknown match	0.17*	0.12
	(0.10)	(0.10)
femaleXmale match	0.18	0.11
	(0.11)	(0.11)
score in Section 1		-0.09***
		(0.01)
risk		-0.01
		(0.01)
Dependent variable mean	0.53	0.53
F-test of female (p-value)	<0.0001	0.0006
F-test of gender of match (p-value)	0.4580	0.7281
Observations	593	590
R-squared	0.06	0.16

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Not payment confident is a dummy variable that equals 1 if the participant thinks his/her payment is below average, and equals 0 otherwise. *Score in Section 1* is an integer between 0 and 8. *F-test of femal*e tests against the null hypothesis that female, femaleXgender-unknown match, and femaleXmale match are jointly zero. *F-test of gender of match* tests against the null hypothesis that male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly zero.

The F-test of female reported at the bottom of the table demonstrates that gender significantly predicts payment confidence. The F-test of gender of match, however, indicates that the gender of the opponent does not affect payment-confidence. We also find that female participants' payment-confidence does not respond differently than men to the gender of their opponent (F-test p-value not shown). Column 2 reveals that, as expected, receiving a higher score in Section 1 significantly increases payment-confidence.

In sum, we find the following results:

1. Men are more willing and more inclined to take risks than women, which is consistent with the literature.
2. Men are more confident than women in their performance *and* payment in the first problem solving section. This again replicates the gender gap found in past studies.

4.3 Gender Gap in Competitiveness

4.3.1 In the absence of feedback

In terms of competitiveness, we first focus on people who did not receive feedback, as past literature has consistently shown a significant gender gap in tournament entry in the absence of performance feedback. We find an economically meaningful (7 percentage points), but not statistically significant gender gap in the no feedback condition (Column 1 of Table 5).

When we allow tournament entry to differ with respect to the gender of the opponent (Column 3 of Table 5), we find that the gender difference emerges when women expect to face a male opponent. F-test of female at the bottom of the table shows that the gender difference in tournament entry is significant ($p < 0.05$) controlling for the gender of the opponent. When facing a female opponent, women are more likely to enter competition than men ($p < 0.1$). However, when facing either a male opponent for sure or when there is a likelihood of facing a male opponent in the gender-unknown condition, women are significantly less likely to enter the tournament. This finding is in line with NV and other previous work that restricts the design to mixed-gender competitions without feedback.

Table 5. Gender Gap in Competitiveness
Dependent variable: tournament entry

Samples	(1) no feedback	(2) forced feedback	(3) no feedback	(4)	(5) forced feedback	(6)	(7) no feedback with non-female opponent	(8) forced feedback with non-female opponent
female	-0.07	0.01	0.22*	0.39**	0.08	-0.08	-0.18**	-0.01
	(0.07)	(0.06)	(0.13)	(0.16)	(0.12)	(0.14)	(0.08)	(0.06)
gender-unknown match			0.12	0.33**	0.04	-0.13		
			(0.12)	(0.16)	(0.10)	(0.13)		
male match			0.17	0.37**	0.05	-0.19		
			(0.14)	(0.18)	(0.12)	(0.14)		
femaleXgender-unknown opponent			-0.39**	-0.57**	-0.03	0.27		
			(0.16)	(0.22)	(0.14)	(0.17)		
femaleXmale opponent			-0.43**	-0.60**	-0.19	0.15		
			(0.19)	(0.25)	(0.16)	(0.18)		
risk				0.04**		0.06***		
				(0.02)		(0.01)		
score-confidence				0.02		-0.01		
				(0.02)		(0.02)		
score in Section 1				0.09		0.09		
				(0.17)		(0.13)		
payment in Section 1				-0.26		-0.05		
				(0.79)		(0.59)		
attribution				0.0034**		0.00096		
				(0.002)		(0.001)		
Dependent variable mean	0.39	0.35	0.39	0.39	0.35	0.35		
F-test of female (p-value)			0.0472	0.0419	0.6117	0.2109		
F-test of gender of match (p-value)			0.1054	0.0736	0.5361	0.2979		
Observations	204	288	204	104	288	189	148	220
R-squared	0.01	0.00	0.04	0.20	0.01	0.17	0.04	0.00

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
F-test of female tests against the null hypothesis that female, femaleXgender-unknown match, and femaleXmale match are jointly zero.
F-test of gender of match tests against the null hypothesis that male match, gender-unkown match, femaleXmale match, and femaleXgender-unknown match are jointly zero.

4.3.2 In the presence of feedback

Studies on the effect of feedback have documented that feedback helps eliminate the gender gap in tournament entry (Ertac and Szentes 2010), so one might expect participants who receive feedback to exhibit no gender difference in competitiveness.[10] To test this, we compare the gender gap in tournament entry for participants who did not receive feedback and those who did – Comparing Columns 1 and 2 of Table 5, we observe that the gap decreases to almost zero in forced feedback without controls for behavior traits and scores (Column 2 of Table 5). Comparing Columns 3 and 4 to Columns 5 and 6, the gender-differential reaction to gender of the opponent disappears when people received feedback. F-test of female at the bottom shows that men and women do not differ in the rate of tournament entry. This finding suggests that feedback, on average, eliminates gender gap in tournament entry.

To further verify the effect of feedback on tournament entry, we restrict the sample to participants who faced a male or gender-unknown opponent and investigate the gender gap in no feedback condition and in forced feedback condition (Columns 7 and 8 of Table 5). While men are significantly more likely to enter competition in the no feedback condition, they do not behave differently from women once provided a feedback. Our findings are consistent with Ertac and Szentes (2010).

4.3.3 Differential selection into tournament by gender

We also investigate how women and men self-select into tournament. Figure 4 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 between tournament entry and score, indicating that low-performing men are more likely to enter competition as their score

[10] In future analysis, we exclude people who are in optional feedback condition for two reasons: 1. Very few participants in optional feedback condition got feedback because participants were unwilling to pay the randomly drawn price. Those who selected into getting feedback may be different from those who received feedback for free. 2. The opportunity of paying for feedback might change participants' reaction and behavior afterward, making the optional feedback condition not comparable to the no feedback condition or the forced feedback condition.

decreases. When feedback on relative payment is present, men and women exhibit similar pattern of tournament entry – higher probability of entering as score increase. Generally, the tournament entry pattern of men seems to be more affected by feedback. Low performing men do not enter competition once they receive feedback on their relative payment.

Figure 4. Local linear regression of tournament entry on score

To verify the selection pattern into tournament, we run the following OLS regression:

$$Tournament\ entry = \beta_0 + \beta_1 female + \beta_2 maleXscore\ in\ Section\ 1 + \beta_3 female\ Xscore\ in\ Section\ 1 + \epsilon,$$

examining participants who did not receive feedback and participants who did separately. We present results in Table 6.

Table 6. Gender Difference in Self-selection into Tournament
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)
Samples		No Feedback			Forced Feedback	
female	-0.07	-0.27*	-0.25	0.01	0.10	0.19
	(0.07)	(0.16)	(0.16)	(0.06)	(0.13)	(0.13)
maleXscore in Section 1		0.01	-0.01		0.07***	0.08***
		(0.03)	(0.03)		(0.02)	(0.02)
femaleXscore 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	0.39	0.35	0.35	0.35
F-test of interactions (p-value)		0.1215	0.0813		0.6114	0.4832
Observations	204	204	204	288	288	287
R-squared	0.01	0.04	0.06	0.0002	0.05	0.13

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
In Columns 1 to 3, we restrict the sample to participants who were assigned to the no feedback condition. In Columns 4 to 6, we restrict the sample to participants who were assigned to the forced feedback condition. F-*test of interactions* tests against the null hypothesis that maleXscore in Section 1 is equal to femaleXscore in Section 1.

In the absence of feedback (Columns 1 to 3), scoring one point higher is associated with an approximately seven percentage points 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 tournament entry pattern. Similar results hold after controlling for risk and confidence. The F-test of interactions at the bottom of the table shows that women and men react differently to an increase in score. These results resemble the findings by Exley, Niederle, and Vesterlund (2016). They observe that women positive self-select into negotiation, whereas men do not. Our results serve as a mirror image in tournament entry. In the absence of feedback, women's self-selection into tournament seems to be more efficient than men's. Men enter competition regardless of how well they perform.

Upon receiving a feedback on relative payment (Columns 4 to 6), men and women do not differ in tournament entry pattern for an additional point in Section 1 (F-test of interactions at the bottom of the table). Both men and women are more likely to enter competition as their score in Section 1 increases.

In short, we find that:

1. Conditional on having faced a male opponent or a gender-unknown opponent, men are more likely than women to enter competition in the absence of feedback.
2. On average, feedback eliminates the observed conditional gender gap in tournament entry.
3. In the absence of feedback, women positively self-select into tournament based on score, whereas men do not. Provided feedback on relative payment, both men and women positively self-select into competition.

4.4 Gender gaps in attribution of feedback

In this section, we investigate whether men and women react differently to feedback. In particular, upon receiving positive and negative feedback, do men and women differently attribute the feedback to their own ability versus luck.

Question 1: Given *negative* feedback, are women more likely than men to attribute it to *lack of ability*?

Question 2: Given *positive* feedback, are women more likely than men to attribute it to *good luck*?

To test if there is a differential attribution of feedback, we adopt the following regression specification:[11]

$$Attribution = \beta_0 + \beta_1\, female + \beta_2\, maleXnegative\ feedback + \beta_3\, femaleXnegative\ feedback + \epsilon,$$

restricting to participants who received feedback.

We separate the analyses into two cases, depending on whether participants are payment-confident. The results are in Table 7.

Among participants who are not payment-confident (Columns 1 and 2 of Table 7), men who receive negative feedback confirming their belief, as opposed to receiving positive feedback, attribute the negative outcome to their lack of ability ($p < 0.05$). Women react similarly, although the adjustment in attribution is smaller than that of men. An F-test of interactions shows that there is no significant difference in attribution between men and women after receiving reinforcing negative feedback.

Among participants who are payment-confident (Columns 3 and 4 of Table 7), men who receive negative feedback attribute the negative surprise to bad luck ($p < 0.01$). On the other hand, women who receive the same, surprising, negative feedback, attribute the negative outcome to their lack of ability ($p < 0.05$). F-test of interactions at the bottom of the table shows that there is a significant difference in attribution between men and women after receiving the negative surprise ($p < 0.01$). Note that the coefficient on female is also significant in this case, indicating that, when receiving positive feedback, payment-confident women are more likely than their male counterparts to attribute it to good luck ($p < 0.05$).

[11] Wave 3 of the experiment does not ask for participant's attribution, so this analysis is only based on Wave 1 and Wave 2.

Table 7. Gender Difference in Attribution of Feedback
Dependent variable: attribution

	(1)	(2)	(3)	(4)
Samples	Not Payment-Confident		Payment-Confident	
female	12.83	11.22	-13.65**	-13.42**
	(10.47)	(10.77)	(5.96)	(6.30)
maleXnegative feedback	28.79***	26.09**	-22.35***	-21.18***
	(9.57)	(10.81)	(5.14)	(7.53)
femaleXnegative feedback	12.65*	10.17	14.06**	15.21*
	(7.05)	(8.21)	(6.55)	(7.92)
score in Section 1		-0.79		0.50
		(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	0.10	0.11	0.21	0.21

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Only Wave 1 and Wave 2 of our experiment included the attribution question. Therefore, this regression analysis is based on these two waves. The entire analysis shown in this table is restricted to participants who received feedback. In Columns 1 and 2, titled "Not Payment-Confident", we further restrict the sample to participants who were not payment-confident. In Columns 3 and 4, titled "Payment-Confident", we restrict the sample to participants who were payment-confident. *F-test of interactions* tests against the null hypothesis that maleXnegative feedback is equal to femaleXnegative feedback.

To verify our findings with the effect of feedback, we consider those who did not get feedback and compare their attribution to those who did get feedback:

$$Attribution = \beta_0 + \beta_1\, female + \beta_2\, maleXforced + \beta_3\, femaleXforced + \epsilon,$$

where $forced$ is a dummy variable that equals 1 when the participant is in the forced feedback condition, and equals 0 if the participant is in the no feedback condition.

We note the following caveat: the measure of attribution differs slightly for people in the no feedback condition and in the forced feedback condition. Consider the following example. If a participant self-evaluated to have an above-average payment, then if she were assigned to no feedback condition, then she would be attributing her *success* to her ability or luck, whereas if she were assigned to forced feedback condition and received a negative feedback, then she would be attributing her *failure* to either her ability or her luck. In short, whether or not a participant receives feedback determines whether she is attributing her *perceived* payment or her *actual* payment, which potentially complicates our interpretation with the interaction of $femaleXforced$ in the above regression. Nonetheless, we expect to see similar pattern with this specification.

We separate the analyses into four cases, depending on participants' payment-confidence and feedback positivity: positive reinforcement, negative reinforcement, positive surprise, and negative surprise. We define the four cases below. Table 8 presents the results.

	Payment-confident	**Not payment-confident**
(Potential) positive feedback[12]	Positive reinforcement	Positive surprise
(Potential) negative feedback	Negative surprise	Negative reinforcement

After getting a reinforcing negative feedback (Columns 1 and 2), both men ($p < 0.1$) and women ($p < 0.1$) attribute it to ability. There is no significant gender difference in the attribution pattern.

[12] Since some participants in this analysis did not receive feedback at all, we say a participant receives potential positive feedback if she either receives positive feedback or would received positive feedback if she were to receive one. Similarly for potential negative feedback.

After receiving a surprising negative feedback (Columns 3 and 4), men attribute the negative surprise to luck, although the change is not significant. Women, on the other hand, attribute the negative outcome to their lack of ability ($p < 0.01$). The F-test at the bottom of the table shows that the change in attribution after receiving a negative surprise is significantly different for men and women ($p < 0.01$).

After getting a reinforcing positive feedback, women ascribe it to luck, whereas men do not update their attribution compared to those who did not receive feedback. The gender difference in attribution is significant ($p < 0.1$).

After getting a surprising positive feedback, both men and women ascribe it more to luck, although the change in not significant for either gender.

To summarize, we find the following differential pattern of attribution of feedback to luck versus ability between men and women:

1. Women attribute negative feedback to lack of ability, regardless of whether it is consistent with their self-evaluation.
2. Men attribute negative feedback to bad lack when the feedback underlines their positive self-evaluation. They attribute negative feedback to lack of ability only when they hold a negative self-evaluation initially.
3. Upon receiving a surprising positive feedback, women attribute it more to luck relative to men.

Table 8. Gender Difference in Attribution of Feedback (including participants who did not receive feedback)
Dependent variable: attribution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Samples	negative reinforcement		negative surprise		positive reinforcement		positive surprise	
female	-0.25	-2.37	-7.57	-7.94	14.00*	18.86**	12.33	5.46
	(7.23)	(7.33)	(9.31)	(9.93)	(8.05)	(9.15)	(17.73)	(15.98)
maleXforced	14.53**	14.53*	-9.00	-9.56	-8.30	-1.99	-15.05	-13.61
	(7.26)	(7.47)	(8.14)	(7.70)	(8.54)	(8.59)	(14.96)	(22.20)
femaleXforced	11.46**	10.71*	21.33***	21.59***	-35.94***	-33.58***	-14.55	-1.48
	(5.57)	(5.66)	(7.38)	(7.47)	(5.38)	(5.35)	(14.56)	(16.28)
score in Section 1		0.34		-0.02		3.44		-2.52
		(1.62)		(1.89)		(2.56)		(7.01)
score-confidence		-0.94		2.30		0.73		7.21**
		(1.54)		(2.59)		(2.30)		(3.08)
risk		-1.31		-0.16		0.48		-1.79
		(0.89)		(1.45)		(1.08)		(2.69)
Dependent variable mean	66.13	66.13	62.44	62.44	73.65	73.65	56.80	56.80
F-test of interactions (p-value)	0.7382	0.6855	0.0073	0.0045	0.0088	0.0011	0.9811	0.5689
Observations	136	135	79	79	49	49	30	30
R-squared	0.06	0.08	0.13	0.15	0.18	0.27	0.10	0.28

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Only Wave 1 and Wave 2 of our experiment included the attribution question. Therefore, this regression analysis is based on these two waves. The entire analysis shown in this table is restricted to participants who were assigned to the no feedback condition or the forced feedback condition. In *negative reinforcement*, we further restrict to participants who were not payment-confident and who would receive negative feedback confirming their belief. In *negative surprise*, we further restrict to participants who were payment-confident and who would receive negative feedback. In *positive reinforcement*, we further restrict to participants who were payment-confident and would receive positive feedback. In *positive surprise*, we further restrict to participants who were not payment-confident and who would receive positive feedback. *F-test of interactions* tests against the null hypothesis that maleXforced is equal to femaleXforced.

4.5 Attribution as an explanatory variable for tournament entry

Our results document a gender gap in tournament entry when the opponent is not female and feedback is unavailable. We also observe a gender difference in attribution of feedback. Provided an unexpected negative feedback, women attribute it to their lack of ability, whereas men attribute it to bad luck. Our next goal is to address if the difference in attribution translates into behavior and thus explains the gender gap in tournament entry. If attribution is a major explanatory variable for the gender gap in tournament entry, then we should expect to see that gender predicts tournament entry in a weaker manner conditional on the attribution of feedback than unconditionally.

In Table 9, we first verify a significant gender gap in tournament entry among participants who faced a male opponent or a gender-unknown opponent *and* who received no feedback (Column 1). Women are 18 percentage points less likely to enter competition than men. Since we want to use attribution as a control, we further restrict the sample to participants who answered the attribution question in Column 2, which provides a baseline level of the gender gap. The gender difference in tournament entry remains in Column 2 and, additionally, in Column 3, where we control for score in Section 1.

In Column 4, we additionally control for attribution and its interaction with gender, since attribution may have different effect on tournament entry for men and women. Upon inclusion of attribution, the coefficient on female is not statistically different from zero. However, the magnitude remains economically meaningful. In Column 5, we further control for risk and payment-confidence, as well as their interactions with gender, allowing them to affect tournament entry differently for men and women. Now, the gender gap in competitiveness drops in magnitude to close to zero. The combined effect of risk, payment-confidence, and attribution seems to eliminate the predicative power of gender on tournament entry.

Table 9. Attribution as an Explanatory Variable for the Gender Gap in Competitiveness
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)
female	-0.18**	-0.27**	-0.22*	-0.23	-0.03
	(0.08)	(0.12)	(0.13)	(0.26)	(0.33)
score in Section 1			0.06*	0.04	0.04
			(0.03)	(0.03)	(0.03)
attribution				0.003	0.004
				(0.003)	(0.004)
femaleXattribution				0.00009	0.00007
				(0.004)	(0.004)
risk					0.04
					(0.04)
femaleXrisk					-0.01
					(0.04)
not payment-confident					0.10
					(0.20)
femaleXnot payment-confident					-0.22
					(0.25)
Dependent variable mean	0.37	0.34	0.34	0.34	0.34
Observations	148	71	71	71	71
R-squared	0.04	0.08	0.12	0.15	0.19

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
For all of the regressions, we restrict to participants who received no feedback, *and* who were not matched with a female opponent. For Columns 2 through 5, we further restrict to participants who were in Wave 1 and Wave 2, since only participants in these two waves answered the attribution question.

However, we cannot determine which ones of the three factors explain the gender gap. Since the regression is restricted to the no feedback condition, feedback attribution, risk aversion, and payment-confidence are all endogenous variables that are not manipulated by treatment. Thus, we cannot disassociate the treatment effects of these variables and make valid inference about which variables are causally meaningful.

We further investigate if attribution is a channel through which *feedback* narrows the gender gap in tournament entry. We found that feedback influences attribution differently for men and women. We now examine whether the predicative power of women and feedback on tournament entry drop as we control for attribution. Table 10 presents the results. We again separate the analyses into two cases, based on participants' payment-confidence. In Columns 1 and 7, we estimate regressions similar to those in Table 7, but using tournament entry as the dependent variable, allowing gender and feedback both to affect tournament entry. We find that women who receive negative feedback, either as reinforcement or as surprise, are deterred from tournament entry. Recall from Table 7 that women who receive negative feedback tend to attribute it more to their own ability, regardless of whether the feedback is consistent with their initial self-evaluation. On the other hand, men's tournament entry pattern generally does not change after they receive a negative feedback. The coefficient is significant for men in Column 7, but the significance soon drops once we control for score in Section 1. In Column 2, we condition on actual score of Section 1, and find that entry deterrence generally persists for women who receive reinforcing negative feedback. We do the same for participants who receive negative surprise in Column 8 and find that entry deterrence vanishes for men and significantly decreases for women, suggesting that actual performance is an explanatory factor.

In Columns 3, 4, 5, we condition additionally on risk aversion and score-confidence for those who receive negative reinforcement. We fail to find evidence showing that these factors are the main channels through which gender and feedback predict entry. Similar results hold in Columns 9, 10, 11 for those who receive negative surprise.

Finally, in Columns 6 and 12, we condition instead on attribution and its interaction with negative feedback, in order to account for the differential interpretations of attribution depending on the direction of feedback. We note that the coefficient changes slightly in Column 6 but dramatically from negative to positive in Column 12. However, the standard errors increase substantially, potentially due to inefficiency of the OLS estimator because of the correlation between covariates. We lack statistical power to conclude definitively that attribution is a channel through which gender and feedback predict tournament entry.

In short, we do not have conclusive evidence supporting attribution as a major explanatory variable of the gender gap in tournament entry. We conclude that:

1. The combination of attribution, risk, and payment-confidence renders the gap statistically and economically insignificant, but we cannot discern the individual effect of the three variables.
2. While we document strong gender differences in attribution of feedback, we are unable to confirm that feedback attribution is a significant channel for the effect of feedback on tournament entry.

Table 10. Attribution as a Channel through which Feedback Influences the Gender Gap in Tournament Entry
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Samples	Not Payment-Confident						Payment-Confident					
female	0.30	0.30	0.38*	0.30	0.36*	0.31	4.75E-17	0.003	0.08	0.01	0.08	0.08
	(0.21)	(0.21)	(0.21)	(0.22)	(0.21)	(0.21)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.14)
maleXnegative feedback	-0.05	-0.08	0.0016	-0.08	0.0014	-0.13	-0.23*	-0.07	-0.07	-0.06	-0.07	0.41
	(0.19)	(0.22)	(0.21)	(0.22)	(0.21)	(0.35)	(0.13)	(0.15)	(0.14)	(0.15)	(0.14)	(0.37)
femaleXnegative feedback	-0.35**	-0.38**	-0.30*	-0.38**	-0.30*	-0.44	-0.38**	-0.23	-0.21	-0.22	-0.21	0.18
	(0.14)	(0.15)	(0.16)	(0.15)	(0.16)	(0.33)	(0.15)	(0.18)	(0.17)	(0.18)	(0.17)	(0.42)
score in Section 1		-0.01	0.01	-0.01	0.01	-0.01		0.08**	0.07**	0.06	0.07*	0.06*
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)		(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
risk			0.05**		0.06**				0.06***		0.06***	
			(0.02)		(0.02)				(0.02)		(0.02)	
score-confidence				0.0008	-0.02					0.02	0.01	
				(0.03)	(0.03)					(0.03)	(0.03)	
attribution						-0.0008						0.01
						(0.004)						(0.004)
attributionXnegative feedback						0.0010						-0.01
						(0.005)						(0.01)
Dependent variable mean	0.30	0.30	0.30	0.30	0.30	0.30	0.41	0.41	0.41	0.41	0.41	0.41
F-test of interactions	0.2107	0.2155	0.2040	0.2179	0.2097	0.1877	0.4680	0.4253	0.4868	0.4297	0.4891	0.2882
Observations	97	97	96	97	96	97	93	93	93	93	93	93
R-squared	0.09	0.09	0.14	0.09	0.14	0.09	0.09	0.14	0.22	0.14	0.22	0.16

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Only Wave 1 and Wave 2 of our experiment included a question about attribution. Therefore, this regression analysis is based on these two waves. The entire analysis shown in this table is restricted to participants who were assigned to forced feedback condition. F-test of interactions tests against the null hypothesis that maleXnegative feedback is equal to femaleXnegative feedback.

Chapter 5. Conclusion

Our findings are generally consistent with the findings of the literature in terms of gender gap in behavioral traits. We find large, significant gender differences in estimation of score, self-evaluation of payment, and self-reported risk level, consistent with the literature. We also find substantial differences in competitiveness by gender among participants who were assigned to mixed-gender groups and received no feedback. Women are significantly less likely to enter competition, which is consistent with the findings by NV, despite the fact that our tournament is relatively less competitive than competition in NV and other previous work. Also consistent with previous literature, we find that the gender gap in tournament entry is eliminated when the participants receive feedback (Ertac and Szentes 2010) and when female participants are made aware that they would be facing a woman in competition (Gneezy, Niederle, and Rustichini 2003).

We present two main findings. First, we observe that women efficiently self-select into tournament based on their performance, even if they have no information about their score or relative performance is provided. Men, on the other hand, blindly enter competition regardless of how they perform in the absence of feedback. Upon receiving feedback, men adjust their tournament entry pattern to the one that is more similar to women – both are more likely to enter as their performance increases. These results resemble the finding by Exley, Niederle, and Vesterlund (2016), who observed that women positively select into negotiations and know when to ask.

Our second main finding contributes to a growing literature on gender differences and reactions to feedback, advancing a potential explanation for the gender difference in competitiveness. Previous work has mostly focused on the direct effect of feedback on decision making, without exploring the channel. We, instead, focuses on the *mechanism* through which feedback influences competition entry. We hypothesize that attribution of feedback is the likely mechanism. We find a gender difference in attribution of feedback to luck versus ability. Women tend to blame ability for negative feedback and credit luck for positive feedback. Men blame ability if negative feedback confirms their self-evaluation but blame bad luck if surprised by negative feedback. They do not react to positive feedback substantially. Our results are largely consistent with our priors, and show that attribution of feedback has a substantial gender difference, which may potentially explain various outcome variables.

We believe this study provides23 opportunities for future work. In this paper, we attempted to explore the underlying reasons for the gender gap in tournament entry, but we did not obtain dispositive evidence. In future work, we plan to explore whether attribution explains tournament entry in a more competitive tournament, such as the winner-take-all tournament scheme in NV. Second, we focused on competitiveness in this study. In future work, we plan to study other outcome variables such as human capital investment since gender differences in attribution could affect investment. Third, we plan to compare our results to gender differences in the impact of feedback and attribution in a stereotypically female task. As noted above, the anagram task that we selected for this comparison was not suited for the online platform since participants could access online search engines.

There are obvious limitations to our experimental framework. First, the experimental setting often involves significant simplification compared to the real world, so the conclusions in this simple world might not hold when there are other factors influencing behavior. Second, the stakes involved in our experiment are low (around two dollars), and thus might not be generalized to high-stakes scenarios. Third, the participants on Amazon Mechanical Turk are usually experienced survey takers. They may exhibit different decision-making traits than the average person.

Despite these limitations, our results have certain implications. Our results on self-selection into tournament show that women correctly sort into competition based on their performance even in the absence of feedback. This finding suggests that it is not always optimal to force women to enter competition, especially when the competition favors men in nature. Our results on feedback attribution highlight a significant disparity: Receiving the same signals about performance, women might not update their beliefs in the same way as men do. This observation has potentially important implications for the labor market: The best feedback for women should emphasize the importance of their ability. Moreover, we may expect attribution of feedback to play an important role in entry or decision-making in the financial markets. Unlucky female traders may misattribute losses to ability rather than luck and are thus deterred from risk-taking. Closing the gender gap in attribution could therefore help to eliminate the gap in investment behavior. However, to draw these real-world conclusions, further empirical investigation is necessary.

References

"Race, Combinations of Two Races, and Not Hispanic or Latino: 2010". 2010 Census Summary File 1. United States Census Bureau. 2010. Retrieved January 2, 2015.

Almenberg, J., and A. Dreber. Gender, stock market participation and financial literacy. *SSE. No. 737*. EFI Working Paper Series, 2012.

Aronson, Elliot, Timothy D. Wilson, and Robin M. Akert. "Social psychology 9th Ed." *New Jersey: Upper Saddle River*(2013).

Bajtelsmit, Vickie L., and Jack L. VanDerhei. "Risk aversion and pension investment choices." *Positioning pensions for the twenty-first century* 45 (1997): 66.

Bannier, Christina E., and Milena Neubert. "Gender differences in financial risk taking: The role of financial literacy and risk tolerance." *Economics Letters* 145 (2016): 130-135.

Barber, Brad M., and Terrance Odean. "Boys will be boys: Gender, overconfidence, and common stock investment." *The quarterly journal of economics* 116, no. 1 (2001): 261-292.

Bengtsson, Claes, Mats Persson, and Peter Willenhag. "Gender and overconfidence." *Economics letters* 86, no. 2 (2005): 199-203.

Berlin, Noémi, and Marie-Pierre Dargnies. "Gender differences in reactions to feedback and willingness to compete." *Journal of Economic Behavior & Organization* 130 (2016): 320-336.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economic*s 2, no. 3 (2010): 228-55.

Beyer, Sylvia, and Edward M. Bowden. "Gender differences in seff-perceptions: Convergent evidence from three measures of accuracy and bias." *Personality and Social Psychology Bulletin* 23, no. 2 (1997): 157-172.

Blau, Francine D., and Lawrence M. Kahn. "Do cognitive test scores explain higher US wage inequality?." *Review of Economics and statistics* 87, no. 1 (2005): 184-193.

Brandts, Jordi, Klarita Gërxhani, and Arthur Schram. *Status-Ranking Aversion*. No. 999. 2017.

Brandts, Jordi, Valeska Groenert, and Christina Rott. "The impact of advice on women's and men's selection into competition." *Management Science* 61, no. 5 (2014): 1018-1035.

Bucher-Koenen, Tabea, Rob Alessie, Annamaria Lusardi, and Maarten Van Rooij. *Women, confidence, and financial literacy*. Mimeo, 2016.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. "Gender, competitiveness, and career choices." *The Quarterly Journal of Economics* 129, no. 3 (2014): 1409-1447.

Byrnes, James P., David C. Miller, and William D. Schafer. "Gender differences in risk taking: A meta-analysis." *Psychological bulletin* 125, no. 3 (1999): 367.

Carpenter, Jeffrey P., Rachel S. Frank, and Emiliano Huet-Vaughn. "Gender Differences in Interpersonal and Intrapersonal Competitive Behavior." (2017).

Charness, Gary, and Uri Gneezy. "Strong evidence for gender differences in risk taking." *Journal of Economic Behavior & Organization* 83, no. 1 (2012): 50-58.

Charness, Gary, Uri Gneezy, and Alex Imas. "Experimental methods: Eliciting risk preferences." *Journal of Economic Behavior & Organization* 87 (2013): 43-51.

Cho, Seo-Young. "Explaining gender differences in confidence and overconfidence in math." (2017).

Comeig, Irene, Alfredo Grau-Grau, Ainhoa Jaramillo-Gutiérrez, and Federico Ramírez. "Gender, self-confidence, sports, and preferences for competition." *Journal of Business Research* 69, no. 4 (2016): 1418-1422.

Crespo, Nuno, Nadia Simoes, and Sandrina B. Moreira. "Gender differences in occupational mobility–evidence from Portugal." *International Review of Applied Economics* 28, no. 4 (2014): 460-481.

Croson, Rachel, and Uri Gneezy. "Gender differences in preferences." *Journal of Economic literature* 47, no. 2 (2009): 448-74.

Current Population Survey, US Census Bureau, 2014.

de Bassa Scheresberg, Carlo. "Financial literacy and financial behavior among young adults: Evidence and implications." *Numeracy* 6, no. 2 (2013): 5.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9, no. 3 (2011): 522-550.

Dwyer, Peggy D., James H. Gilkeson, and John A. List. "Gender differences in revealed risk taking: evidence from mutual fund investors." *Economics Letters* 76, no. 2 (2002): 151-158.

Eckel, Catherine C., and Philip J. Grossman. "Rebate versus matching: does how we subsidize charitable contributions matter?."*Journal of Public Economics* 87, no. 3-4 (2003): 681-701.

Eckel, Catherine C., and Sascha C. Füllbrunn. "Thar she blows? Gender, competition, and bubbles in experimental asset markets." *American Economic Review* 105, no. 2 (2015): 906-20.

Eil, David, and Justin M. Rao. "The good news-bad news effect: asymmetric processing of objective information about yourself." *American Economic Journal: Microeconomics* 3, no. 2 (2011): 114-38.

Ertac, Seda, and Balazs Szentes. "The effect of performance feedback on gender differences in competitiveness: experimental evidence." *Work. Pap., Koc Univ., Turkey* (2010).

Exley, Christine L., Muriel Niederle, and Lise Vesterlund. Knowing when to ask: The cost of leaning in. No. w22961. National Bureau of Economic Research, 2016.

Feng, Jing, Ian Spence, and Jay Pratt. "Playing an action video game reduces gender differences in spatial cognition." *Psychological science* 18, no. 10 (2007): 850-855.

García-Mainar, Inmaculada, Guillermo García-Martín, and Victor M. Montuenga. "Gender differences in occupations, job attributes, and job satisfaction." *Revista de Economía Aplicada* 24, no. 71 (2016).

Gentile, Monica, Nadia Linciano, and Paola Soccorso. "Financial Advice Seeking, Financial Knowledge and Overconfidence. Evidence from the Italian Market." (2016).

Germine, Laura, Ken Nakayama, Bradley C. Duchaine, Christopher F. Chabris, Garga Chatterjee, and Jeremy B. Wilmer. "Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments." *Psychonomic bulletin & review* 19, no. 5 (2012): 847-857.

Gillen, Ben, Erik Snowberg, and Leeat Yariv. *Experimenting with measurement error: Techniques with applications to the caltechcohort study*. No. w21517. National Bureau of Economic Research, 2015.

Gino, Francesca, Alison Wood Brooks, and Maurice E. Schweitzer. "Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice." *Journal of personality and social psychology* 102, no. 3 (2012): 497.

Gneezy, Uri, and Aldo Rustichini. "Gender and competition at a young age." *American Economic Review* 94, no. 2 (2004): 377-381.

Gneezy, Uri, Kenneth L. Leonard, and John A. List. "Gender differences in competition: Evidence from a matrilineal and a patriarchal society." *Econometrica* 77, no. 5 (2009): 1637-1664.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. "Performance in competitive environments: Gender differences." *The Quarterly Journal of Economics* 118, no. 3 (2003): 1049-1074.

Goldin, Claudia, and Lawrence F. Katz. "Transitions: Career and family life cycles of the educational elite." *American Economic Review* 98, no. 2 (2008): 363-69.

Goldin, Claudia. "A grand gender convergence: Its last chapter." *American Economic Review* 104, no. 4 (2014): 1091-1119.

Goldin, Claudia. "Claudia Goldin: Will More of Our Daughters Grow up to Be Economists?" *Ohio.com*. October 17, 2013. Accessed March 1, 2017.

Grosse, Niels, Gerhard Riener, and Markus Dertwinkel-Kalt. "Explaining gender differences in competitiveness: Testing a theory on gender-task stereotypes." (2014).

Halko, Marja-Liisa, Markku Kaustia, and Elias Alanko. "The gender effect in risky asset holdings." *Journal of Economic Behavior & Organization* 83, no. 1 (2012): 66-81.

Healy, Andrew, and Jennifer Pate. "Can teams help to close the gender competition gap?." *The Economic Journal* 121, no. 555 (2011): 1192-1204.

Hirshleifer, David, Angie Low, and Siew Hong Teoh. "Are overconfident CEOs better innovators?." *The Journal of Finance* 67, no. 4 (2012): 1457-1498.

Hoffmann, Arvid OI, and Thomas Post. "Self-attribution bias in consumer financial decision-making: How investment returns affect individuals' belief in skill." *Journal of Behavioral and Experimental Economics* 52 (2014): 23-28.

Holt, Charles A., and Susan K. Laury. "Risk aversion and incentive effects." *American economic review* 92, no. 5 (2002): 1644-1655.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein. "Social interaction and stock-market participation." *The journal of finance* 59, no. 1 (2004): 137-163.

Iriberri, Nagore, and Pedro Rey-Biel. Let's (not) talk about sex: Gender awareness and stereotype-threat on performance under competition. Barcelona GSE Working Papers Series, Department of Economics, Universitat Pompeu Fabra. http://www. ed. ac. uk/polopoly_fs/1.106296, 2013.

Jeworrek, Sabrina. "Competition Entry and Relative Performance Feedback: The Importance of Information Disaggregated by Gender." (2016).

Jianakoplos, Nancy Ammon, and Alexandra Bernasek. "Are women more risk averse?." *Economic inquiry* 36, no. 4 (1998): 620-630.

Jung, Seeun, Chung Choe, and Ronald L. Oaxaca. "Gender Wage Gaps and Risky vs. Secure Employment: An Experimental Analysis." (2016).

Kagel, John H., and Alvin E. Roth, eds. *The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics*. Princeton university press, 2016.

Kamas, Linda, and Anne Preston. "The importance of being confident; gender, career choice, and willingness to compete." *Journal of Economic Behavior & Organization* 83, no. 1 (2012): 82-97.

Kiss, Hubert J., Ismael Rodriguez-Lara, and Alfonso Rosa-Garcia. "Do women panic more than men? An experimental study of financial decisions." *Journal of Behavioral and Experimental Economics* 52 (2014): 40-51.

Kleinjans, Kristin J. "Do gender differences in preferences for competition matter for occupational expectations?." *Journal of Economic*

Koellinger, Philipp, Maria Minniti, and Christian Schade. "Gender differences in entrepreneurial propensity." *Oxford bulletin of economics and statistics* 75, no. 2 (2013): 213-234.

Korsgaard, M. Audrey. "The impact of self-appraisals on reactions to feedback from others: The role of self-enhancement and self-consistency concerns." *Journal of Organizational Behavior* (1996): 301-311.

Kramer, Marc M. "Financial literacy, confidence and financial advice seeking." *Journal of Economic Behavior & Organization*131 (2016): 198-217.

Liden, Robert C., and Terence R. Mitchell. "Reactions to feedback: The role of attributions." *Academy of Management Journal* 28, no. 2 (1985): 291-308.

Love, E. Geoffrey, Denice Williams Love, and Gregory B. Northcraft. "Is the end in sight? Student regulation of in-class and extra-credit effort in response to performance feedback." *Academy of Management Learning & Education* 9, no. 1 (2010): 81-97.

Lusardi, Annamaria, and Olivia S. Mitchell. "Planning and financial literacy: How do women fare?." *American Economic Review* 98, no. 2 (2008): 413-17.

Masters, M. S. and B. Sanders. "Is the Gender Difference in Mental Rotation Disappearing?" *Behavior Genetics* 23, no. 4 (1993): 337.

McCarty, Paulette A. "Effects of feedback on the self-confidence of men and women." *Academy of Management Journal* 29, no. 4 (1986): 840-847.

Mezulis, Amy H., Lyn Y. Abramson, Janet S. Hyde, and Benjamin L. Hankin. "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, no. 5 (2004): 711.

Miller, Dale T., and Michael Ross. "Self-serving biases in the attribution of causality: Fact or fiction?." *Psychological bulletin*82, no. 2 (1975): 213.

Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. “Managing self-confidence: Theory and experimental evidence.” No. w17014. National Bureau of Economic Research, 2011.

Niederle, Muriel, and Lise Vesterlund. "Do women shy away from competition? Do men compete too much?." *The Quarterly Journal of Economics* 122, no. 3 (2007): 1067-1101.

Niederle, Muriel, and Lise Vesterlund. "Gender and competition." *Annu. Rev. Econ.* 3, no. 1 (2011): 601-630.

Niederle, Muriel. “Gender.” Chapter 8 in *The Handbook of Experimental Economics 2* edited by John Kagel and Alvin E. Roth, pp. 481 563, Princeton University Press, Princeton, NJ, 2016.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. "Running experiments on amazon mechanical turk." (2010).

Paolcci, Gabriele, Jesse Chandler, and Panagiotis Ipeirotis. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making* 5 (5): 411-9.

Peters, Michael, and Christian Battista. "Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library." *Brain and* cognition 66, no. 3 (2008): 260-264.

Prast, Henriette, Mariacristina Rossi, Costanza Torricelli, and Dario Sansone. "Do women prefer pink? The effect of a gender stereotypical stock portfolio on investing decisions." *Politica economica* 31, no. 3 (2015): 377-420.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. "Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender." *The Economic Journal* 127, no. 604 (2017): 2153-2186.

Reuben, Ernesto, Pedro Rey-Biel, Paola Sapienza, and Luigi Zingales. "The emergence of male leadership in competitive environments." *Journal of Economic Behavior & Organization*83, no. 1 (2012): 111-117.

Richler, Jennifer J., Jeremy B. Wilmer, and Isabel Gauthier. "General object recognition is specific: Evidence from novel and familiar objects." *Cognition* 166 (2017): 42-55.

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

Saccardo, Silvia, Aniela Pietrasz, and Uri Gneezy. "On the size of the gender difference in competitiveness." *Management Science* (2017).

Sarsons, Heather. "Gender differences in recognition for group work." Harvard University Working Paper (2015).

Shurchkov, Olga, and Catherine C. Eckel. "Gender Differences in Behavioral Traits and Labor Market Outcomes." in *The Oxford Handbook of Women and the Economy* edited by Susan L. Averett, Laura M. Argys, and Saul D. Hoffman: Oxford University Press, 2017.

Shurchkov, Olga. "Under pressure: gender differences in output quality and quantity under competition and time constraints." *Journal of the European Economic Association*10, no. 5 (2012): 1189-1213.

Sunden, Annika E., and Brian J. Surette. "Gender differences in the allocation of assets in retirement savings plans." *The Ameican Economic Review* 88, no. 2 (1998): 207-211.

Van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie. "Financial literacy and stock market participation." *Journal of Financial Economics* 101, no. 2 (2011): 449-472.

van Veldhuizen, Roel. Gender differences in tournament choices: Risk preferences, overconfidence or competitiveness?. No. 14. Discussion Paper, 2017.

Vandenberg, Steven G., and Allan R. Kuse. "Mental rotations, a group test of three-dimensional spatial visualization." *Perceptual and motor skills* 47, no. 2 (1978): 599-604.

Weeks, Justin W., Tejal A. Jakatdar, and Richard G. Heimberg. "Comparing and contrasting fears of positive and negative evaluaion as facets of social anxiety." *Journal of Social and Clinical Psychology* 29, no. 1 (2010): 68-94.

Wolfer, Justin. "Even Famous Female Economists Get No Respect." The New York Times, November 11, 2015. Accessed March 1, 2018.

Appendix 1. Pilot Study and Full-length Experiment Design

1. Pilot study

We conducted two rounds of pilot study:

In the first round, participants worked on three separate tasks, each for 5 minutes. The three tasks are Mental Rotation Test (MRT), Finding Median Task (FMT), and Matrix Pattern Task (MPT).[13] Examples of these tasks are shown below.

After each section of tasks, participants answered a few questions:

1. Confidence perception: "How many questions do you think you solved correctly in this task?"
2. Gender perception: "Do you think men or women are more likely to get a higher score in this task?"

Participants were paid a piece rate payment based on how many questions they solved correctly in the three tasks.

We collected 49 valid answers for the first round of pilot study. A Mann-Whitney U test cannot reject the null hypothesis that men and women performed identically in all three tasks ($p = 0.1276$ for MRT; $p = 0.2580$ for FMT; $p = 0.3357$ for MPT). In terms of gender perception associated with the test, only MRT is perceived to favor men ($p = 0.0306$ for MRT; $p = 0.8881$ for FMT; $p = 0.2017$ for MPT).

In the second round of pilot study, participants completed the MRT and MPT. FMT was dropped in this round of pilot because it does not seem to be associated with any gender perception. Participants answered the same confidence and gender perception questions. The only difference is that they were paid by a tournament-based payment scheme that is the same as in the main study. In particular, we say:

"We previously ran this test on a similar group and got the performance. We will randomly pair you with a participant from that group. Your payment will be as follows:

1. If your total score is higher than your random match's, you get 20 cents for each correct answer.
2. If your total score is lower than your random match's, you get 15 cents for each correct answer."

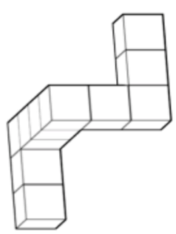
Their score was compared to the score of a randomly chosen participant from the first round of pilot, and their payment was calculated as stated above.

We collected 50 valid answers for the second round of pilot study. In terms of performance, men scored higher than women in MPT ($p = 0.0657$). The performance is similar across gender in MRT ($p = 0.3983$). Neither task was associated with a strong gender perception ($p = 0.4016$ for MPT; $p = 0.3319$ for MRT). A summary table is shown below. Taken together, MRT is perceived as favoring male and does not elicit gender gap in performance; therefore, we decided to use MRT for our main experiment.

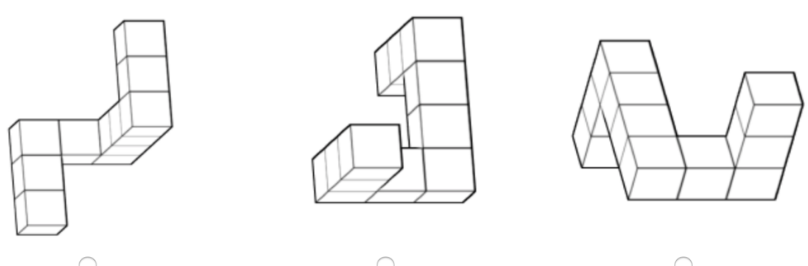
[13] The version of MRT adopted in our experiment is modified on Peters and Battista (2008)'s Mental Rotation Stimulus Library. The MPT adopted in our experiment is a selection of an online test by Richler, Wilmer, and Gauthier (2017). Both tests are provided to us by Prof. Jeremy Wilmer.

Round 1: Piece rate Payment		
	Difference in Performance	Difference in Perception
MRT	N	Y
FMT	N	N
MPT	N	N
Round 2: Tournament Payment		
	Difference in Performance	Difference in Perception
MRT	N	N
MPT	Y	N

Sample Mental Rotation Test (MRT) question:



Select the shape that is a rotated version of the one above:



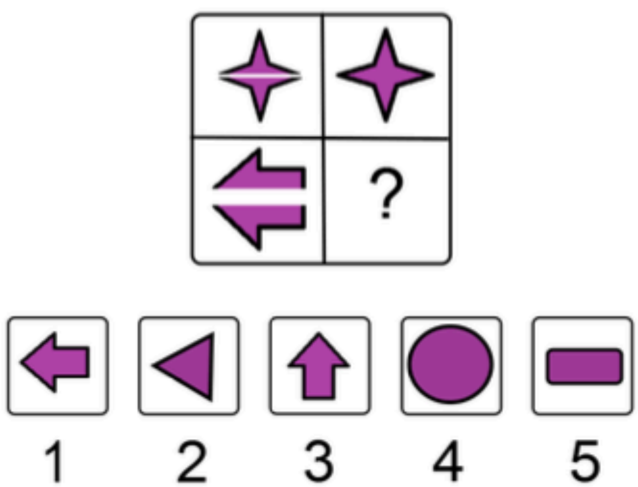
Sample Finding Median Test (FMT) question:

Select the median of these 9 numbers.

2.34	0.16	3.52
4.58	7.14	2.86
1.64	8.27	6.82

Sample Matrix Pattern Test (MPT) question:

In each problem, you will see a sequence of shapes. Your job is to fill in the question mark with a shape from the choices.



2. Full-length experimental design:

Round 1:

1. Participants compete a 2.5-minute MRT
2. Confidence evaluation: "*How many problems do you think you solved correctly?*"
3. Payment information:
   "*We previously ran this test on a similar group and got the performance. We will randomly pair you with a participant from that group. Your payment will be as follows:*
   - *If your total score is higher than your random match's, you get 20 cents for each correct answer.*
   - *If your total score is lower than your random match's, you get 15 cents for each correct answer.*"
4. Participants learn their payment. Participants are randomly assigned into two conditions in this step:
   a. Gender-unknown condition
      "*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.*"
   b. Gendered condition
      "*Our matching process has randomly matched you with a male participant from the other group. Your score has been compared with his, and your payment is shown below.*"
      OR
      "*Our matching process has randomly matched you with a female participant from the other group. Your score has been compared with hers, and your payment is shown below.*"
5. Self-evaluation: "*Do you think it's an above average payment or below average payment?*"
6. Participants are randomly assigned into one of the three feedback conditions:
   a. No Feedback condition
      i. Participants assign relative importance to luck and ability.
   b. Forced feedback condition
      i. Participants learn whether their payment is above or below average:
         "*We compared your payment to the average payment of the other group. Your payment is above (below) the average.*"
      ii. Then they assign relative importance to luck and ability.
   c. Optional feedback condition
      i. Participants have an opportunity to choose whether they want feedback about whether their payment is above or below average:
         "*Now you have an opportunity to find out if your payment is above the average payment of the other group. The knowledge of your **relative payment** will inform your decision in the next round and influence your payment.*
         *Please select the **maximum amount** you are willing to pay to receive this information. Please note that, if you select a certain amount, we automatically assume you are willing to pay any price that is lower than*

your selection. For example, if you select 10 cents, we automatically assume that you are also willing to pay 5 cents.
Then, we will randomly draw a price from these five options. If this price is less than or equal to your maximum, then the price will be subtracted from your payment, and you will receive information about your relative payment. On the other hand, if this price is greater than your maximum, then you will not get any information, and your payment will be unaffected.

Please note that, in a rare instance, your price of information may be greater than your final bonus payment, in which case, the difference will be deducted from your base payment.

- 5 cents
- 10 cents
- 15 cents
- 20 cents
- 25 cents"

ii. Then we randomly draw one of the amount. If one indicates that she would be willing to pay that amount, she will pay it and receive the feedback information. If one indicates that she would not be willing to pay that amount, she will not pay any money and will not receive any information.

iii. Finally, participants assign relative importance to luck and ability

Round 2

1. Participants choose which payment scheme they want to use for the second round:
   i. **Option 1**: We randomly pair you with another participant from the other group. I
      - If your total score is higher than your randomly chosen match's, you get 25 cents for each correct answer.
      - If your total score is lower than your randomly chosen match's, you get 10 cents for each correct answer.

   ii. **Option 2**: You get 17.5 cents for each correct answer, regardless of anyone else's score.
2. Participants complete a 2.5-minute test. They payment is calculated based on the payment scheme they choose.

In the end, participants complete a short demographic questionnaire.

Demographic questionnaire:

Please rank from 1-10 how you see yourself:
Are you generally a person who avoid taking risks or are you fully prepared to take risks?
(1 is unwilling to take risks and 10 is fully prepared to take risks)

Do you think men or women are more likely to get a high score in this task?

- Men
- Women

What is your gender?

a. Male
b. Female
c. Other/Do not wish to disclose

What is your age in years?

What is the highest level of education you have completed?

a. Less than high school
b. High school or GED
c. Some college
d. 2-year college degree (Associates)
e. 4-year college degree (BA, BS)
f. Master's degree (MA, MS)
g. Doctoral degree (PhD)
h. Professional degree (MD, JD, DDS, etc.)

Are you of Hispanic origin or descent, such as Mexican, Puerto Rican, Cuban, or other Spanish background?

a. Yes
b. No

Which of the following best describes your race?

a. White
b. African-American or Black
c. Asian
d. Native Hawaiian or Other Pacific Islander
e. Native American
f. Other/Do not wish to disclose

Which of the following best describes your annual household income before taxes?

a. Less than $10,000
b. $10,000 - $19,999
c. $20,000 - $29,999
d. $30,000 - $39,999
e. $40,000 - $49,999
f. $50,000 - $74,999
g. $75,000 - $99,999
h. $100,000 - $149,999
i. $150,000 - $249,999
j. $250,000-$499,999
k. $500,000 and over

Appendix 2. Analysis of Wave 1

In this wave, participants completed the full-length experiment. Our analysis is based on 307 valid responses.

Table 1. Gender Gap in Risk Preference, Score-confidence, Payment-confidence, and Tournament Entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dependent variables	risk	score-confidence	score-confidence	not payment-confident	not payment-confident	tournament entry	tournament entry
female	-1.24***	-0.87***	-0.62***	0.25***	0.19***	-0.09	-0.05
	(0.27)	(0.21)	(0.20)	(0.06)	(0.05)	(0.05)	(0.05)
score in Section 1			0.35***		-0.09***		0.05***
			(0.05)		(0.01)		(0.02)
Dependent variable mean	4.86	3.93	3.93	0.57	0.57	0.31	0.31
Observations	304	307	307	307	307	307	307
R-squared	0.07	0.06	0.17	0.06	0.16	0.01	0.04

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
In Column 1, risk is on a scale from 1 to 10, with 0 representing extreme risk aversion and 10 representing extreme willingness to take risks. In Columns 2 and 3, score-confidence is a measured on a scale of 0 to 10. As a comparison, the highest possible score in Section 1 is 8. In Columns 6 and 7, tournament entry equals 1 if participants choose the tournament-based payment scheme and equals 0 otherwise.

Table 2. Influence of Gender of the Opponent on Payment-confidence
Dependent variable: not payment-confident

	(1)	(2)
female	0.05	0.03
	(0.12)	(0.12)
gender-unknown match	-0.15	-0.14
	(0.11)	(0.10)
male match	-0.15	-0.08
	(0.12)	(0.12)
femaleXgender-unknown match	0.24*	0.19
	(0.14)	(0.14)
femaleXmale match	0.32**	0.19
	(0.16)	(0.16)
score in Section 1		-0.08***
		(0.01)
risk		-0.01
		(0.01)
Dependent variable mean	0.57	0.57
F-test of female (p-value)	<0.0001	0.0128
F-test of gender of match (p-value)	0.2737	0.4755
Observations	307	304
R-squared	0.08	0.17

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Not payment-confident is a dummy variable that equals 1 if the participant thinks his/her payment is below average, and equals 0 otherwise. F-test of female tests if female, femaleXgender-unknown match, and femaleXmale match are jointly differently from zero. F-test of gender of match tests if male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly different from zero.

Table 3. Gender Gap in Competitiveness
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Samples	no feedback	forced feedback	no feedback		forced feedback		no feedback with non-female opponent	forced feedback with non-female opponent
female	-0.11	-0.08	0.23	0.39**	-0.26	-0.28	-0.27**	-0.04
	(0.10)	(0.09)	(0.17)	(0.16)	(0.22)	(0.21)	(0.11)	(0.10)
gender-unknown match			0.20	0.33**	-0.08	-0.16		
			(0.17)	(0.16)	(0.21)	(0.20)		
male match			0.28	0.37**	-0.13	-0.23		
			(0.19)	(0.18)	(0.24)	(0.22)		
femaleXgender-unknown match			-0.50**	-0.57**	0.23	0.36		
			(0.22)	(0.22)	(0.26)	(0.24)		
femaleXmale match			-0.49*	-0.60**	0.23	0.46*		
			(0.26)	(0.25)	(0.29)	(0.27)		
risk				0.04**		0.08***		
				(0.02)		(0.02)		
score-confidence				0.02		-0.01		
				(0.02)		(0.03)		
score in Section 1				0.09		0.02		
				(0.17)		(0.16)		
payment in Section 1				-0.26		0.16		
				(0.79)		(0.73)		
attribution				0.003**		0.0002		
				(0.002)		(0.002)		
Dependent variable mean	0.36	0.30	0.36	0.36	0.30	0.30	0.34	0.31
F-test of female (p-value)			0.084	0.0419	0.6852	0.3529		
F-test of gender of match (p-value)			0.1662	0.0736	0.8586	0.4368		
Observations	104	102	104	104	102	101	71	83
R-squared	0.01	0.01	0.08	0.20	0.02	0.18	0.08	0.00

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
F-test of female tests against the null hypothesis that female, femaleXgender-unknown match, and femaleXmale match are jointly zero. F-test of gender of match tests against the null hypothesis that male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly zero.

Table 4. Gender Difference in Self-selection into Tournament
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)
Samples	No feedback condition			Forced feedback condition		
female	-0.11	-0.39*	-0.36	-0.08	0.01	0.04
	(0.10)	(0.22)	(0.22)	(0.09)	(0.24)	(0.22)
maleXscore in Section 1		0.02	0.01		0.04	0.04
		(0.04)	(0.04)		(0.04)	(0.04)
femaleXscore in Section 1		0.11***	0.10***		0.02	0.04
		(0.03)	(0.03)		(0.04)	(0.04)
risk			0.03			0.08***
			(0.02)			(0.02)
score-confidence			0.02			-0.01
			(0.02)			(0.03)
Dependent variable mean	0.36	0.36	0.36		0.3	0.3
F-test of interactions (p-value)		0.1029	0.0827		0.4872	0.3173
Observations	104	104	104	102	102	101
R-squared	0.01	0.09	0.12	0.01	0.02	0.15

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
In Columns 1 to 3, we restrict the sample to participants who were assigned to the no feedback condition. In Columns 4 to 6, we restrict the sample to participants who were assigned to the forced feedback condition. F-test of interactions tests against the null hypothesis that maleXscore in Section 1 is equal to femaleXscore in Section 1.

Table 5. Gender Difference in Attribution of Feedback
Dependent variable: attribution

	(1)	(2)	(3)	(4)
Samples	Not Payment-Confident		Payment-Confident	
female	-5.69	-5.71	-12.80	-14.31
	(4.83)	(13.72)	(9.97)	(10.65)
maleXnegative feedback	-0.07	-0.19	-25.40***	-36.75***
	(5.37)	(13.10)	(5.92)	(11.73)
femaleXnegative feedback			9.53	-1.92
			(10.46)	(14.45)
score in Section 1		-0.31		-1.87
		(2.71)		(2.45)
score-confidence		0.41		-3.45
		(2.47)		(2.60)
risk		-0.79		0.61
		(1.85)		(1.53)
Dependent variable mean	71.08	71.08	65.88	65.88
F-test of interactions (p-value)			0.0003	0.0019
Observations	51	50	51	51
R-squared	0.01	0.01	0.21	0.28

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
The entire analysis shown in this table is restricted to participants who received feedback. In Columns 1 and 2, titled "Not Payment-Confident", we further restrict the sample to participants who were not payment-confident. In Column 3 and 4, titled "Payment Confident", we restrict the sample to participants who were payment-confident. F-test of interactions tests against the null hypothesis that maleXnegative feedback is equal to femaleXnegative feedback. No female participants in this wave received positive feedback when they were not payment confident, so the interaction of female and negative feedback is omitted in Columns 1 and 2.

Table 6. Gender Difference in Attribution of Feedback (including participants who did not receive feedback)
Dependent variable: attribution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Samples	negative reinforcement		negative surprise		positive reinforcement		positive surprise	
female	-0.25	-2.16	-7.57	-7.59	14.00	29.38**	12.33	-3.78
	(7.25)	(7.37)	(9.33)	(10.02)	(9.01)	(10.70)	(21.84)	(35.58)
maleXforced	15.38*	14.63*	-8.95	-10.47	-5.20	-8.13	14.67	29.08
	(8.09)	(8.72)	(8.49)	(7.64)	(9.69)	(10.00)	(14.56)	(8.22)
femaleXforced	10.01	9.84	20.75***	21.04***	-32.00**	-55.90***		
	(6.16)	(6.25)	(7.67)	(7.78)	(10.60)	(13.31)		
score in Section 1		0.34		0.00		23.28**		-3.68
		(1.74)		(1.97)		(7.52)		(36.38)
score-confidence		-0.21		2.47		-4.70		4.01
		(1.66)		(2.85)		(5.10)		(2.98)
risk		-1.45		0.03		3.65		4.77
		(1.01)		(1.58)		(3.17)		(2.05)
Dependent variable mean	66.13	66.13	62.44	62.44	73.65	73.65	56.80	56.80
F-test of interactions (p-value)	0.5986	0.6590	0.0116	0.0051	0.0889	0.0063		
Observations	113	112	71	71	15	15	7	7
R-squared	0.05	0.07	0.12	0.14	0.26	0.64	0.11	0.74

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
The entire analysis shown in this table is restricted to participants who were assigned to the no feedback condition or the forced feedback condition. In negative reinforcement, we further restrict to participants who were not payment-confident and who would receive negative feedback confirming their belief. In negative surprise, we further restrict to participants who were payment-confident and who would receive negative feedback. In positive reinforcement, we further restrict to participants who were payment-confident and would receive positive feedback. In positive surprise, we further restrict to participants who were not payment-confident and who would receive positive feedback. F-test of interactions tests against the null hypothesis that maleXforced is equal to femaleXforced. No female participants in this wave received positive feedback when they held a were not payment-confident, so the interaction of female and forced is omitted in Columns 7 and 8.

Table 7. Attribution as an Explanatory Variable for the Gender Gap in Competitiveness
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)
female	-0.27**	-0.22*	-0.23	-0.03
	(0.12)	(0.13)	(0.26)	(0.33)
score in Section 1		0.06*	0.04	0.04
		(0.03)	(0.03)	(0.03)
attribution			0.00	0.00
			(0.00)	(0.00)
femaleXattribution			0.00	0.00
			(0.00)	(0.00)
risk				0.04
				(0.04)
femaleXrisk				-0.01
				(0.04)
not payment-confident				0.10
				(0.20)
femaleXnot payment-confident				-0.22
				(0.25)
Dependent variable mean	0.34	0.34	0.34	0.34
Observations	148	71	71	71
R-squared	0.04	0.12	0.15	0.19

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
This analysis restricts to participants who received no feedback *and* who were not matched with a female opponent.

Appendix 3. Analysis of Wave 2

In this wave, we corrected the mistake made in wave 1 and included only the forced feedback condition. Our analysis is based on 88 valid responses.

Table 1. Gender Gap in Risk Preference, Score-confidence, Payment-confidence, and Tournament Entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dependent variables	risk	score-confidence	score-confidence	not payment-confident	not payment-confident	tournament entry	tournament entry
female	-1.67***	-0.75*	-0.58	0.30***	0.27***	0.06	0.09
	(0.47)	(0.42)	(0.36)	(0.10)	(0.10)	(0.11)	(0.10)
score in Section 1			0.61***		-0.11***		0.11***
			(0.09)		(0.02)		(0.03)
Dependent variable mean	4.86	3.93	3.93	0.57	0.57	0.31	0.31
Observations	88	88	88	88	88	88	88
R-squared	0.13	0.04	0.33	0.09	0.24	0.00	0.16

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
In Column 1, risk is on a scale from 1 to 10, with 0 representing extreme risk aversion and 10 representing extreme willingness to take risks. In Columns 2 and 3, score-confidence is a measured on a scale of 0 to 10. As a comparison, the highest possible score in Section 1 is 8. In Columns 6 and 7, tournament entry equals 1 if participants choose the tournament-based payment scheme and equals 0 otherwise.

Table 2. Influence of Gender of the Opponent on Payment-confidence
Dependent variable: not payment-confident

	(1)	(2)
female	0.25	0.23
	(0.22)	(0.19)
gender-unknown match	-0.08	-0.05
	(0.20)	(0.17)
male match	-0.08	0.05
	(0.23)	(0.19)
femaleXgender-unknown match	0.13	0.15
	(0.26)	(0.23)
femaleXmale match	-0.07	-0.13
	(0.31)	(0.27)
score in Section 1		-0.11***
		(0.02)
risk		0.0042
		(0.02)
Dependent variable mean	0.53	0.53
F-test of female (p-value)	0.0379	0.0563
F-test of gender of match (p-value)	0.8700	0.8389
Observations	88	88
R-squared	0.10	0.26

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Not payment-confident is a dummy variable that equals 1 if the participant thinks his/her payment is below average, and equals 0 otherwise. F-test of female tests if female, femaleXgender-unknown match, and femaleXmale match are jointly differently from zero. F-test of gender of match tests if male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly different from zero.

Table 3. Gender Gap in Competitiveness
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)
Samples		forced feedback condition		forced feedback with non-female
female	0.06	0.02	0.15	0.07
	(0.11)	(0.22)	(0.19)	(0.12)
gender-unknown match		-0.08	-0.11	
		(0.20)	(0.18)	
male match		-0.08	-0.19	
		(0.23)	(0.18)	
femaleXgender-unknown match		0.20	0.20	
		(0.27)	(0.24)	
femaleXmale match		-0.20	-0.25	
		(0.30)	(0.24)	
risk			0.06***	
			(0.02)	
score-confidence			-0.05	
			(0.03)	
score in Section 1			0.35**	
			(0.16)	
payment in Section 1			-0.94	
			(0.78)	
attribution			0.003	
			(0.002)	
Dependent variable mean	0.41	0.41	0.41	0.39
F-test of female (p-value)		0.4089	0.0908	
F-test of gender of match (p-value)		0.2115	0.0100	
Observations	88	88	88	66
R-squared	0.004	0.06	0.32	0.01

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
F-test of female tests against the null hypothesis that female, femaleXgender-unknown match, and femaleXmale match are jointly zero. F-test of gender of match tests against the null hypothesis that male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly zero.

Table 4. Gender Difference in Attribution of Feedback
Dependent variable: attribution

	(1)	(2)	(3)	(4)
Samples	Not Payment-Confident		Payment-Confident	
female	17.78	16.11	-14.92**	-13.28*
	(10.60)	(11.79)	(7.34)	(7.41)
maleXnegative feedback	32.58***	20.01	-21.83**	-9.27
	(10.66)	(18.44)	(9.90)	(13.10)
femaleXnegative feedback	16.88*	1.67	20.42***	27.53***
	(9.35)	(22.01)	(6.76)	(9.68)
score in Section 1		-4.91		1.96
		(7.10)		(3.15)
score-confidence		0.65		2.36
		(3.20)		(2.52)
risk		-1.35		0.03
		(2.00)		(1.29)
Dependent variable mean	63.74	63.74	69.88	69.88
F-test of interactions (p-value)	0.0041	0.3373	0.0026	0.0043
Observations	46	46	42	42
R-squared	0.22	0.24	0.20	0.30

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
This entire analysis is restricted to those who received feedback. In Columns 1 and 2, titled "Not Payment-Confident", we further restrict the sample to participants who were not payment-confident. In Columns 3 and 4, titled "Payment-Confident", we restrict the sample to participants who were payment-confident. F-test of interactions tests against the null hypothesis that maleXnegative feedback is equal to femaleXnegative feedback.

Appendix 4. Analysis of Wave 3

In this wave, participants completed the full-length experiment. Our analysis is based on 198 valid responses.

Table 1. Gender Gap in Risk Preference, Score-confidence, Payment-confidence, and Tournament Entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dependent variables	risk	score-confidence	score-confidence	not payment-confident	not payment-confident	tournament entry	tournament entry
female	-0.96***	-0.57**	-0.32	0.15**	0.10	0.03	0.04
	(0.33)	(0.28)	(0.26)	(0.07)	(0.07)	(0.07)	(0.07)
score in Section 1			0.50***		-0.09***		0.03
			(0.07)		(0.02)		(0.02)
Dependent variable mean	5.43	3.92	3.92	0.48	0.48	0.39	0.39
Observations	198	198	198	198	198	198	198
R-squared	0.04	0.02	0.23	0.02	0.14	0.0008	0.01

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
In Column 1, risk is on a scale from 1 to 10, with 0 representing extreme risk aversion and 10 representing extreme willingness to take risks. In Columns 2 and 3, score-confidence is a measured on a scale of 0 to 10. As a comparison, the highest possible score in Section 1 is 8. In Columns 6 and 7, tournament entry equals 1 if participants choose the tournament-based payment scheme and equals 0 otherwise.

Table 2. Influence of Gender of the Opponent on Payment-confidence
Dependent variable: not payment confident

	(1)	(2)
female	0.08	0.07
	(0.14)	(0.14)
gender-unknown match	-0.07	-0.01
	(0.12)	(0.12)
male match	-0.08	-0.03
	(0.14)	(0.13)
femaleXgender-unknown match	0.10	0.02
	(0.18)	(0.17)
femaleXmale match	0.08	0.07
	(0.20)	(0.19)
score in Section 1		-0.09***
		(0.02)
risk		0.003
		(0.02)
Dependent variable mean	0.48	0.48
F-test of female (p-value)	0.1899	0.5107
F-test of gender of match (p-value)	0.9688	0.9963
Observations	198	198
R-squared	0.03	0.14

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Not payment-confident is a dummy variable that equals 1 if the participant thinks his/her payment is below average, and equals 0 otherwise. F-test of female tests if female, femaleXgender-unknown match, and femaleXmale match are jointly differently from zero. F-test of gender of match tests if male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly different from zero.

Table 3. Gender Gap in Competitiveness
Dependent variable: tounrnament entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Samples	no feedback	forced feedback	no feedback		forced feedback		no feedback with non-female opponent	forced feedback with non-female opponent
female	-0.03	0.08	0.22	0.24	0.36*	0.45**	-0.10	-0.03
	(0.10)	(0.10)	(0.21)	(0.21)	(0.18)	(0.18)	(0.11)	(0.12)
gender-unknown match			0.03	0.02	0.18	0.20		
			(0.17)	(0.18)	(0.15)	(0.16)		
male match			0.04	0.06	0.21	0.26		
			(0.21)	(0.21)	(0.16)	(0.16)		
femaleXgender-unknown match			-0.28	-0.22	-0.35	-0.41*		
			(0.25)	(0.26)	(0.24)	(0.24)		
femaleXmale match			-0.39	-0.38	-0.46*	-0.46*		
			(0.29)	(0.29)	(0.27)	(0.25)		
risk				0.01		0.06***		
				(0.02)		(0.02)		
score-confidence				0.04		-0.03		
				(0.04)		(0.03)		
score in Section 1				0.06		-0.08		
				(0.17)		(0.18)		
pay in Section 1				-0.31		0.61		
				(0.78)		(0.82)		
Dependent variable mean	0.43	0.36	0.43	0.43	0.36	0.36	0.4	0.4
F-test of female (p-value)			0.5782	0.607	0.2672	0.0982		
F-test of gender of match (p-value)			0.5036	0.6141	0.4703	0.3703		
Observations	100	98	100	100	98	98	77	71
R-squared	0.00007	0.01	0.03	0.06	0.04	0.14	0.01	0.00007

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
F-test of female tests against the null hypothesis that female, femaleXgender-unknown match, and femaleXmale match are jointly zero. F-test of gender of match tests against the null hypothesis that male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly zero.

Table 4. Gender Difference in Self-selection into Tournament
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)
Samples	No feedback condition			Forced feedback condition		
female	-0.03	-0.11	-0.08	0.08	-0.04	0.04
	(0.10)	(0.23)	(0.23)	(0.10)	(0.21)	(0.21)
maleXscore in Section 1		0.002	-0.02		0.03	0.05
		(0.04)	(0.04)		(0.03)	(0.03)
femaleXscore in Section 1		0.03	0.01		0.06*	0.08
		(0.04)	(0.04)		(0.04)	(0.05)
risk			0.01			0.05**
			(0.02)			(0.02)
score-confidence			0.05			-0.04
			(0.03)			(0.03)
Dependent variable mean	0.43	0.43	0.43	0.36	0.36	0.36
F-test of interactions (p-value)		0.8057	0.8182		0.1748	0.1109
Observations	100	100	100	98	98	98
R-squared	0.00	0.00	0.03	0.01	0.04	0.10

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
In Columns 1 to 3, we restrict the sample to participants who are assigned to the no feedback condition. In Columns 4 to 6, we restrict the sample to participants who are assigned to the forced feedback condition. F-test of interactions tests against the null hypothesis that maleXscore in Section 1 is equal to femaleXscore in Section 1.

Appendix 5. Robustness Check

Table 1. Gender Gap in Risk Preference, Score-confidence, Payment-confidence

	(1)	(2)	(3)	(4)
dependent variables	Risk	Score-confidence	Score-confidence	Proportion not payment-confident
female	-1.27***	-0.52***	0.19***	0.004
	(0.19)	(0.15)	(0.04)	(0.04)
score in Section 1		0.44***	-0.09***	0.05***
		(0.04)	(0.01)	(0.01)
Dependent variable mean	5.12	3.94	0.53	0.36
Observations	590	590	590	590
R-squared	0.13	0.24	0.18	0.07

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
All regressions control for race, income, and education. All specifications are based on data from all three waves. In Column 1, risk is on a scale from 1 to 10, with 0 representing extreme unwillingness to take risks and 10 representing extreme willingness to take risks. In Column 2 and 3, score-confidence is measured on a scale of 0 to 10. As a comparison, the highest possible score in Section 1 is 8

Table 2. Influence of Gender of the Opponent on Payment-confidence
Dependent variable: not payment-confident

	(1)	(2)
female	0.10	0.09
	(0.08)	(0.08)
gender-unknown match	-0.13*	-0.09
	(0.07)	(0.07)
male match	-0.12	-0.06
	(0.08)	(0.08)
femaleXgender-unknown match	0.18*	0.12
	(0.10)	(0.10)
femaleXmale match	0.17	0.10
	(0.11)	(0.11)
score in Section 1		-0.09***
		(0.01)
risk		-0.01
		(0.01)
Dependent variable mean	0.53	0.53
F-test of female (p-value)	<0.0001	0.0003
F-test of gender of match (p-value)	0.3963	0.7349
Observations	590	590
R-squared	0.09	0.18

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
All regressions control for race, income, and education. *Not payment confident* is a dummy variable that equals 1 if the participant thinks his/her payment is below average, and equals 0 otherwise. *Score in Section 1* is an integer between 0 and 8. F-test of female tests against the null hypothesis that female, femaleXgender-unknown match, and femaleXmale match are jointly zero. F-test of gender of match tests against the null hypothesis that male match, gender-unknown match, femaleXmale match, and femaleXgender-unknown match are jointly zero

Table 3. Gender Gap in Competitiveness
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Samples	no feedback	forced feedback	no feedback		forced feedback		no feedback with non-female opponent	forced feedback with non-female opponent
female	-0.10	0.02	0.19	0.22	0.07	-0.11	-0.18**	0.01
	(0.07)	(0.06)	(0.14)	(0.18)	(0.12)	(0.16)	(0.08)	(0.07)
gender-unknown match			0.04	0.14	0.05	-0.16		
			(0.13)	(0.19)	(0.10)	(0.13)		
male match			0.11	0.22	0.05	-0.21		
			-0.37**	-0.39	-0.03	0.29		
femaleXgender-unknown opponent			(0.17)	(0.25)	(0.15)	(0.18)		
			(0.22)	(0.22)	(0.26)	(0.24)		
femaleXmale opponent			-0.39*	-0.56*	-0.16	0.16		
			(0.20)	(0.28)	(0.17)	(0.19)		
risk				0.04		0.06***		
				(0.02)		(0.02)		
score-confidence				0.01		-0.01		
				(0.02)		(0.02)		
score in Section 1				0.09		0.04		
				(0.15)		(0.13)		
payment in Section 1				-0.25		0.15		
				(0.72)		(0.63)		
attribution				0.01***		0.001		
				(0.002)		(0.001)		
Dependent variable mean	0.39	0.36	0.39	0.39	0.36	0.36	0.37	0.36
F-test of female (p-value)			0.0785	0.1994	0.7553	0.2444		
F-test of gender of match (p-value)			0.0621	0.2942	0.7392	0.3277		
Observations	204	287	204	104	287	189	148	219
R-squared	0.11	0.05	0.15	0.45	0.05	0.23	0.15	0.06

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
All regressions control for race, income, and education. F-test of female tests against the null hypothesis that female, femaleXgender-unknown match, and femaleXmale match are jointly zero. F-test of gender of match tests against the null hypothesis that male match, gender-unkown match, femaleXmale match, and femaleXgender-unknown match are jointly zero.

Table 4. Gender Difference in Self-selection into Tournament
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)
Samples	No feedback condition			Forced feedback condition		
female	-0.10	-0.25	-0.22	0.02	0.14	0.23
	(0.07)	(0.17)	(0.17)	(0.06)	(0.15)	(0.14)
maleXscore in Section 1		0.01	-0.0006		0.07***	0.08***
		(0.03)	(0.03)		(0.02)	(0.02)
femaleXscore in Section 1		0.06**	0.05		0.05**	0.06**
		(0.03)	(0.03)		(0.02)	(0.02)
risk			0.02			0.06***
			(0.02)			(0.01)
score-confidence			0.03			-0.02
			(0.02)			(0.02)
Dependent variable mean	0.39	0.39	0.39	0.36	0.36	0.36
F-test of interactions (p-value)		0.1109	0.2665		0.0006	0.0003
Observations	204	204	204	287	287	287
R-squared	0.11	0.13	0.14	0.05	0.10	0.17

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
All regressions control for race, income, and education. In Columns 1 to 3, we restrict the sample to participants who were assigned to the no feedback condition. In Columns 4 to 6, we restrict the sample to participants who were assigned to the forced feedback condition. F-test of interactions tests against the null hypothesis that maleXscore in Section 1 is equal to femaleXscore in Section 1.

Table 5. Gender Difference in Attribution of Feedback
Dependent variable: attribution

	(1)	(2)	(3)	(4)
Samples	Not Payment-Confident		Payment-Confident	
female	8.25	7.96	-13.87**	-13.74*
	(14.73)	(15.72)	(6.95)	(7.12)
maleXnegative feedback	27.03*	26.45*	-25.23***	-27.11***
	(13.62)	(14.76)	(6.89)	(8.97)
femaleXnegative feedback	17.52**	16.37*	10.99	10.05
	(7.79)	(8.86)	(8.40)	(9.53)
score in Section 1		-0.36		-1.02
		(2.41)		(2.01)
score-confidence		0.73		0.42
		(2.11)		(1.96)
risk		-0.99		0.39
		(1.46)		(1.07)
Dependent variable mean	67.60	67.60	67.69	67.69
F-test of interactions (p-value)	0.0107	0.0575	0.0012	0.0018
Observations	96	96	93	93
R-squared	0.23	0.23	0.34	0.34

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
All regressions control for race, income, and education. Only Wave 1 and Wave 2 of our experiment included the attribution question. Therefore, this regression analysis is based on these two waves. The entire analysis shown in this table is restricted to participants who received feedback. In Columns 1 and 2, titled "Not Payment-Confident", we further restrict the sample to participants who were not payment-confident. In Columns 3 and 4, titled "Payment-Confident", we restrict the sample to participants who were payment-confident. F-test of interactions tests against the null hypothesis that maleXnegative feedback is equal to femaleXnegative feedback.

Table 6. Gender Difference in Attribution of Feedback (including participants who did not receive feedback)
Dependent variable: attribution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Samples	negative reinforcement		negative surprise		positive reinforcement		positive surprise	
female	7.09	5.39	-12.33	-14.72	21.80*	31.13**	22.49	19.92
	(7.72)	(8.21)	(8.76)	(8.93)	(12.64)	(13.82)	(30.93)	(31.06)
maleXforced	21.67***	21.21**	-17.82*	-18.44*	-7.89	10.68	22.36	19.75
	(7.92)	(8.41)	(10.21)	(9.74)	(9.76)	(12.13)	(28.70)	(33.24)
femaleXforced	9.77*	9.32	18.82**	19.05**	-45.01***	-37.25***	-16.76	-5.04
	(5.72)	(5.74)	(7.75)	(7.64)	(9.73)	(8.90)	(20.79)	(47.88)
score in Section 1		0.23		-0.09		2.76		-0.13
		(1.65)		(2.34)		(2.77)		(18.24)
score-confidence		-1.11		0.92		2.84		5.27
		(1.77)		(3.02)		(2.70)		(6.70)
risk		-0.50		-1.35		-0.18		1.99
		(0.95)		(1.53)		(1.47)		(5.57)
Dependent variable mean	65.99	65.99	62.44	62.44	73.65	73.65	56.80	56.80
F-test of interactions (p-va	0.0054	0.0110	0.0073	0.0036	0.0001	0.0009	0.5761	0.7572
Observations	135	135	79	79	49	49	30	30
R-squared	0.23	0.24	0.32	0.33	0.39	0.48	0.53	0.66

Notes: Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
All regressions control for race, income, and education. Only Wave 1 and Wave 2 of our experiment included the attribution question. Therefore, this regression analysis is based on these two waves. The entire analysis shown in this table is restricted to participants who were assigned to the no feedback condition or the forced feedback condition. In negative reinforcement, we further restrict to participants who were not payment-confident and who would receive negative feedback confirming their belief. In negative surprise, we further restrict to participants who were payment-confident and who would receive negative feedback. In positive reinforcement, we further restrict to participants who were payment-confident and would receive positive feedback. In positive surprise, we further restrict to participants who were not payment-confident and who would receive positive feedback. F-test of interactions tests against the null hypothesis that maleXforced is equal to femaleXforced.

Table 7. Attribution as an Explanatory Variable for the Gender Gap in Competitiveness
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)
female	-0.18**	-0.32**	-0.29**	-0.32	-0.09
	(0.09)	(0.12)	(0.13)	(0.29)	(0.54)
score in Section 1			0.04	0.03	0.03
			(0.03)	(0.03)	(0.04)
attribution				0.01*	0.01*
				(0.003)	(0.004)
femaleXattribution				0.0002	-0.0005
				(0.004)	(0.005)
risk					0.01
					(0.06)
femaleXrisk					-0.02
					(0.06)
not payment-confident					0.12
					(0.24)
femaleXnot payment-confident					-0.17
					(0.30)
Dependent variable mean	0.37	0.34	0.34	0.34	0.34
Observations	148	71	71	71	71
R-squared	0.15	0.40	0.43	0.50	0.51

Notes: Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
All regressions control for race, income, and education. For all of the regressions, we restrict to participants who received no feedback, and who were not matched with a female opponent. For Columns 2 through 5, we further restrict to participants who were in Wave 1 and Wave 2, since only participants in these two waves answered the attribution question.

Table 8. Attribution as a Channel through which Feedback Influences the Gender Gap in Tournament Entry
Dependent variable: tournament entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Samples	Not Payment-Confident						Payment-Confident					
female	0.25	0.23	0.29	0.24	0.28	0.24	-0.01	0.0005	0.05	-0.01	0.05	0.08
	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)	(0.17)	(0.17)	(0.16)	(0.17)	(0.17)	(0.16)
maleXnegative feedback	-0.08	-0.15	-0.10	-0.14	-0.10	-0.26	-0.26	-0.07	-0.08	-0.08	-0.09	0.30
	(0.21)	(0.24)	(0.24)	(0.25)	(0.24)	(0.37)	(0.16)	(0.19)	(0.18)	(0.20)	(0.18)	(0.40)
femaleXnegative feedback	-0.36**	-0.42**	-0.37**	-0.42**	-0.37**	-0.55	-0.41**	-0.23	-0.18	-0.21	-0.17	0.01
	(0.16)	(0.17)	(0.18)	(0.17)	(0.18)	(0.35)	(0.19)	(0.21)	(0.20)	(0.22)	(0.20)	(0.47)
score in Section 1		-0.03	-0.02	-0.03	-0.01	-0.03		0.08**	0.07*	0.07	0.07	0.08*
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
risk			0.04*		0.04				0.07***		0.07***	
			(0.02)		(0.03)				(0.02)		(0.02)	
score-confidence				0.01	-0.01					0.03	0.01	
				(0.03)	(0.04)					(0.04)	(0.03)	
attribution						-0.0008						0.01
						(0.005)						(0.004)
attributionXnegative feedback						0.002						-0.004
						(0.01)						(0.01)
Dependent variable mean	0.30	0.30	0.30	0.30	0.30	0.30	0.41	0.41	0.41	0.41	0.41	0.41
F-test of interactions	0.0726	0.0602	0.1375	0.0599	0.1401	0.2651	0.0408	0.5568	0.6593	0.6229	0.6839	0.4194
Observations	96	96	96	96	96	96	93	93	93	93	93	93
R-squared	0.24	0.25	0.27	0.25	0.28	0.25	0.20	0.25	0.32	0.25	0.32	0.27

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
The entire analysis shown in this table is restricted to participants who were in Wave 1 or Wave 2 and were assigned to forced feedback condition. All regressions control for race, income, and education. In Column 1 through 6, titled "Negative Self-Evaluation", we further restrict the sample to participants who were not payment-confident. In Column 7 through 12, titled "Positive Self-Evaluation", we restrict the sample to participants who were payment-confident. F-test of interactions tests against the null hypothesis that maleXnegative feedback is equal to femaleXnegative feedback.