Economic Mobility in Tiered Competitions: On Gender Differences in Competitive Sorting[*]

Valon Vitaku[†] Marco A. Palma[‡] Brian A. Toney[§]

February 8, 2025

Abstract

We introduce competition tiers to the seminal study of Niederle and Vesterlund (2007) over repeated periods to investigate competitive sorting behavior in a dynamic setting. In this new competitive environment, there is a gender gap in willingness to compete driven by women *Skipping-the-Top* most competitive tier and men *Skipping-the-Bottom* non-competitive tier. Despite receiving feedback that largely corrects beliefs about past relative performance, gender differences in competitive choices persist unless opportunity costs of alternative payment schemes are disclosed independent of compensation choices. Emotional expressions suggest medium and low ability men struggle accepting downward mobility in competitive choices is efficient.

JEL classification: C90, D83, J16, J24, J33

Keywords: experiment, facial expressions, feedback, gender, mobility, tournament entry

1 Introduction

The seminal study by Niederle and Vesterlund (2007), henceforth NV, revealed a pronounced gender difference in preferences for competing. Since then, an extensive literature on this topic has emerged reproducing a robust finding that men enter competitions at higher rates than women of similar ability (Markowsky & Beblo, 2022). Previous experimental work on this topic concentrates on binary tournament entry decisions; to the best of our knowledge, we are the first to introduce multi-tiered competitive brackets in repeated interactions to investigate the sorting efficiency of compensation choices. Different job occupations typically have multiple tiers, each with varying degrees of responsibility, as well as differences in wages and prestige. Upward economic mobility is an important human aspiration and a key economic indicator for opportunity and equality measures. The question of whether there is a gender disparity in upward mobility is important, as it may help explain gender inequities in compensation and job promotion. In this study, we use the laboratory to control for ability and other factors that are difficult to disentangle in naturally occurring environments. In the experiment, the Top Competitive bracket (Tier I) carries a large reward for the best player, while the Bottom Competitive bracket (Tier II) provides a modest reward for above median performance. A non-competitive bracket is also available providing a piece rate compensation scheme. Participants are randomly assigned to groups of four and there is an efficient allocation for the best player to select the top Tier I, the second best to select Tier II and the remaining two players to opt for the non-competitive (piece-rate) bracket.

At the start of the repeated game, we find a gender gap in willingness to compete driven by women *Skipping-the-Top* most competitive bracket and men *Skipping-the-Bottom* non-competitive bracket. This initial finding closely relates to recent work by Fluchtmann, Glenny, Harmon, and Maibom (2024) who document a 'gender application gap' where women tend to choose lower paying jobs even after accounting for ability and other observable characteristics. Despite receiving feedback that largely corrects beliefs about past relative performance, our findings suggest that the gender gap in competitive choices remains through the end of the ten-period game unless opportunity costs of alternative payment schemes are fully disclosed regardless of compensation choices. This difference is driven by persistent entry of men into more competitive than optimal payment schemes which causes them to incur significant welfare loss. Our study contributes to the existing literature by highlighting dynamics of economic mobility in competitive choices and a persistent gender gap in the context of limited information about opportunity costs of alternative

(non-chosen) compensation schemes. Additionally, to better understand gender differences in competitive choices, we measure emotions using a facial expression analysis software called AFFDEX by Affectiva using noninvasive cameras that were mounted on individual computer terminals (McDuff et al., 2016; Stöckli, Schulte-Mecklenbeck, Borer, & Samson, 2018). In general, the emotions data suggest that women regulate the emotions produced by the competition entry feedback, and men appear to struggle with acceptance that downward mobility in competitive choices is efficient.

The gender gap in competitiveness has a tendency to shrink in teams (Dargnies, 2012; Healy & Pate, 2011; Kuhn & Villeval, 2011) and matrilineal societies (Andersen, Ertac, Gneezy, List, & Maximiano, 2013; Gneezy, Leonard, & List, 2009), while it increases when completing math tasks (Dasgupta, Mani, Sharma, & Singhal, 2019; Dreber, Von Essen, & Ranehill, 2014; Große & Riener, 2010; Shurchkov, 2012), and paradoxically in countries with higher gender equality (Markowsky & Beblo, 2022). Previous literature further suggests that confidence and risk aversion reduce but do not eliminate the gender gap in competitive choices (Cueva, Iturbe-Ormaetxe, Ponti, & Tomás, 2019; Kang, Lei, Song, & Zhang, 2024; Tungodden & Willén, 2023). Feedback aversion is another possible factor that could potentially explain the gender gap in competition. If women are more feedback-averse than men, this could have profound implications for the labor market. For instance, if high-performing women are averse to feedback in a multi-tiered competitive environment, they may never learn that moving up the ladder is revenue maximizing and significant individual and societal gains would be forgone. If this is the case, policies and work environments that encourage transparency and provide information on performance thresholds for different positions could lead to more efficient sorting in labor markets. This investigation is crucial as it touches on the broader question of decision-making dynamics in competitive settings, which has significant implications for labor economics and organizational behavior.

A consequence of avoiding feedback is that people may deprive themselves of potentially useful information that could be utilized to improve decision-making or increase earnings. For instance, Möbius, Niederle, Niehaus, and Rosenblat (2022) find women are significantly more averse to feedback which could be consistent with maintaining self-confidence if women tend to place more weight on anticipatory utility. In implementing the game studying repeated competitive choices, we exogenously vary whether subjects learn the opportunity costs of alternative payment schemes. In the *Conditional-Disclosure* condition, subjects do not learn the performance thresholds of competitive tournaments they needed to surpass if they elected to not compete; instead, the performance threshold of a competitive tournament that subjects needed to beat is provided only if subjects

chose that tournament and after competing for its earnings. In the treatment condition, subjects learn the performance thresholds of competitive tournaments they needed to surpass regardless of their compensation choices. We expect such unconditional disclosure to eliminate the gender gap in willingness to compete and overconfidence as subjects no longer have instrumental or non-instrumental motivations to avoid competitions (Golman, Hagmann, & Loewenstein, 2017; Golman & Loewenstein, 2018). In other words, we expect that information about forgone earnings, or opportunity costs of compensation choices, will prod subjects towards payment schemes that present Pareto improvements if they are not efficiently sorted following the initial compensation choice.

Building on the understanding of how multidimensional sorting influences market outcomes, it is essential to study competition over repeated interactions, especially considering how firm dynamics and on-the-job search significantly affect labor market fluctuations and misallocation (Elsby & Gottfries, 2022). Related, Dohmen and Falk (2011) provide evidence that individual productivity, risk attitudes, and gender significantly influence self-selection into different payment schemes, demonstrating the complex nature of multidimensional sorting in labor markets. In fact, it is beneficial for labor markets to be dynamic for worker mobility (Donovan, Lu, & Schoellman, 2023; Hahn, Hyatt, & Janicki, 2021; Topel & Ward, 1992). However, systemic inequality exists, partially, because workers self-select into roles (Eeckhout, 2018). In particular, women and men have a tendency to self-select into different roles (Flory, Leibbrandt, & List, 2015; Fluchtmann et al., 2024), leading to literature investigating the effect of information in correcting for any internalities (Ewers, 2012). In an attempt to remedy this issue, firms can offer more transparent disclosures about promotions within and across different positions to better align people's skill set with different jobs and their requirements. Such disclosures could include performance expectations or productivity targets that can be objectively measured.

In settings featuring repeated competitive choices, a gender difference in persistence emerges after losing or missing performance targets. Ellison and Swanson (2023) and Kang et al. (2024) find a more pronounced discouragement effect for women than men after losing – among students falling just short of target scores, females are significantly more likely to drop out than males. Wasserman (2023) finds a similar gender difference in persistence among politicians who lost an election which caused over 50% more attrition among female than male candidates. Similarly, Buser and Yuan (2019) find losing a competition negatively and disproportionately influences women's subsequent willingness to compete. Alnamlah and Gravert (2020), however, find women are as likely to keep competing after losing and learning about their performance. Their findings further suggest an

3

asymmetric gender reaction to the cause of losing in competitive settings: women are more likely than men to keep competing if their loss is attributed to a lack of luck while men are more likely to compete than women if their loss is attributed to a lack of ability. Both of these scenarios carry important welfare implications.

Women may also be more reluctant to enter competitions because of how losses might impact their future productivity. Losing is detrimental to productivity for women when luck plays a role whereas for men, a loss impacts negatively on productivity only when the prize at stake is large enough (Gill & Prowse, 2014). Along a similar vein, empirical evidence of high school students receiving information about relative exam performance suggest similar patterns (Goulas & Megalokonomou, 2021). Students in cohorts that observed their peers' scores and who learned they were low achieving compared to their peers performed worse the next time they took the exam, compared to cohorts that did not learn their relative position. Importantly, the results are more pronounced for females indicating greater sensitivity to feedback. Given these findings, women may anticipate that losses would negatively affect their future productivity and opt for competitive schemes less frequently. In light of these implications, women may be more inclined to opt for competitive compensations if they are not required to perform following feedback – hence, women may be more likely to compete toward the end of the experiment. The literature on gender disparities in willingness to compete is abundant and it continues to grow demonstrating its significance in economics and labor markets. However, to date we know little about how women and men elect to sort into competitive tiers under the information disclosures we implement over repeated interactions. We fill this gap in the literature.

Closely Related Literature. Numerous studies investigate the effectiveness of interventions designed to reduce the gender gap in competitiveness while showing there are no gender differences in performance (see the summary by Markowsky and Beblo (2022) and the references therein). Some of these studies provide performance feedback, advice or other information to subjects to reduce this gap. Ertac and Szentes (2011) closely follow the design implemented by NV and provide information on the target score to beat prior to the tournament entry decision. They find that this simple intervention practically eliminates the gender gap in willingness to compete, and the effect is primarily driven by an increased willingness to compete of the high-performing women. Along the same vein, Shastry, Shurchkov, and Xia (2020) find that feedback about one's performance relative to the average of a pool of candidates against whom one is to be matched eliminates this gender gap but the effect is primarily driven by low-performing men opting out of

the tournament. Similarly, Wozniak, Harbaugh, and Mayr (2014) show that relative performance feedback eliminates the gender gap in competitions by providing subjects with an ordered list of the performance of all the participants in their session, with their own performance highlighted, before they chose a compensation scheme. The elimination of the gap in competitiveness, in this study too, is primarily driven by low-ability men opting for less competitive forms of compensation with women's competitive choices remaining virtually unchanged following feedback.

An alternative to contemporaneous sessions is to match subjects with scores of previous subjects from historical sessions (Cason, Masters, & Sheremeta, 2010). Using this method, Cason et al. (2010) provide subjects with the actual piece-rate scores of previous subjects against whom they would compete. This intervention eliminated the gender gap in competitiveness in winner-take-all tournaments but not in proportional-prize contests where males were more likely to enter competitions. Berlin and Dargnies (2016) on the other hand find that providing subjects with feedback on relative performance to the median does not affect the gender gap in competitiveness. While men do not significantly update their future entry decisions for competing as low-performing men expect favorable competition in the future, above (below) median women choose to increase (decrease) their entry rates.

Advice is an additional informational intervention that has shown the ability to reduce the gender gap in willingness to compete. Kessel, Mollerstrom, and Van Veldhuizen (2021) deploy simple informational interventions that highlight evidence supporting the existence of a gender gap in willingness to compete due to differences in competitiveness, risk-taking and confidence. They find that all three interventions reduce the gender gap in willingness to compete. Brandts, Groenert, and Rott (2015) on the other hand show the advice of subjects who previously decided whether to compete improves the entry decisions of subjects, in that forgone earnings due to inefficient entry decisions go significantly down. The improvements are mainly driven by increased entry of strong-performing women and reduced entry of weak-performing men though the overall gender gap persisted due to an emerging gap among intermediate performers.

The rest of the paper is organized as follows: the next section covers the implementation of the experiment, in section 3 we present our results and we conclude in section 4.

2 Experimental Design

The experiment was computerized using o-Tree (Chen, Schonger, & Wickens, 2016). At the beginning of each session, the experimenter read out loud general instructions to establish common knowledge. The interaction between subjects was anonymous. A total of 244 undergraduate students participated in 26 gender-balanced sessions where 2 or 3 groups of 4 subjects participated in each session. Gender was never mentioned to participants though they could observe the (balanced) session composition since they were seated in the same waiting room before entering the lab. Data were collected between October 2023 and February 2024 at the Human Behavior Lab at Texas A&M University. A session lasted, on average, about one hour. Average earnings were $18.61 (sd=6.727), which includes a $10 show-up fee.

2.1 Tasks and Experimental Treatments

The main task of the experiment consists of adding up sets of five two-digit numbers for a three-minute period. We selected this mathematical computation task because it requires both skill and effort, and previous literature has robustly documented there are no gender differences in performance for this task (Markowsky & Beblo, 2022). The computation numbers were randomly drawn at the subject-trial level and subjects could not proceed to the next problem until they submitted a correct answer.[1] Participants were not allowed to use a calculator though scratch paper was provided if they wished to use it. A record tracking the number of correct answers was displayed on the screen during the task.

Participants completed twelve rounds and they were informed that only one of the twelve rounds would be randomly chosen for payment at the end of the experiment. Paying for only one round reduces concerns that decisions in a given round may be used to hedge against outcomes in other rounds. Participants had three minutes in each round to solve as many problems as they could. Given our interest in repeated competitive choices following feedback, we chose the duration of the addition task to be three minutes to obtain sufficient variation in performance while keeping the duration of the experiment feasible. The specific compensation schemes for each round were as follows:

Round 1 – Piece-rate: Subjects are given the three-minute addition task. If Round 1 is randomly selected for payment, they receive 50 cents per correct answer.

---

[1] If a wrong answer was submitted, a message appeared that prompted subjects to correct the calculation. We enforced this rule to prevent subjects from strategically skipping difficult problems for easier ones.

Round 2 – Single-Tier Tournament: Subjects are given the three-minute addition task. If Round 2 is randomly selected for payment, the subject who solves the largest number of correct problems in the group receives $2 per correct answer while the other subjects receive no payment. In case of ties, the winner is chosen randomly among the high scorers.

Rounds 3-12 – Choice of Compensation Scheme for Future Performance: At the beginning of each round and before performing the three-minute addition task, participants select one of the three following payment schemes:

- Tier I

  A subject receives a large reward of $2 per correct answer if her score exceeds that of the other group members from the previous round they just completed; otherwise she receives no payment.

- Tier II

  A subject receives a modest reward of $1 per correct answer if her score is above median (i.e., exceeds the score of at least two other group members) from the previous round they just completed; otherwise she receives no payment.

- Piece-rate

  A subject receives 50 cents per correct answer.

While participants knew their own performance on a task, they were not informed about their relative performance during the first two rounds. At the beginning of each session, groups of 4 were randomly formed and randomly assigned to one of two feedback conditions. Group compositions remained fixed throughout the experiment given our interest in feedback as a mechanism for prodding subjects toward payoff-maximizing payment schemes. The two feedback conditions were introduced at the conclusion of Round 3 and deployed for the remaining rounds:

Conditional-Disclosure: Subjects learned the performance threshold needed to surpass for a given tier only if they chose it and after the competition. Tier scores were not disclosed to subjects who chose the *piece-rate* scheme.

Unconditional-Disclosure: Tier scores were disclosed to subjects independent of their compensation choices.

Screenshot examples from the feedback stage of the two treatments can be found in the Appendix (see Figures A4 and A5) along with the experimental instructions.

7

Winners of tiers I and II are determined based on the relative comparison to all the other group members' previous round performance. This design feature originally adopted by NV presents several distinct advantages. First, the performance of a subject who enters one of the two competitive tiers initially in Round 3 is evaluated against the performance of participants who performed under a tournament compensation. Similarly, in Round 4 and beyond, the performance of a subject who enters a competitive tier is evaluated against the past performance of all the other participants. Second, this design feature eliminates concerns over beliefs regarding the compensation choices of others. In other words, this feature eliminates a potential source of error through biased beliefs about other participants' choices. Further, since a participant's compensation choice does not affect the payment of any other participant, we can rule out other-regarding preferences for the possibility of shying away from competitive environments because winning imposes a negative externality on others.[2] Hence, subjects effectively face individual decision problems in rounds 3-12 which depend on their ability to surpass different performance thresholds from previous rounds and their preference for performing in competitive environments that differ in competition intensity. While this design feature allows for the possibility that all participants in a group can win in *Tier I* if everyone increases their performance in Round 3 or other rounds beyond the highest performance of other group members from the previous round, we expect these occurrences to be unlikely.[3]

We elicited beliefs by asking subjects to guess their rank (between 1 and 4) in the group to account for potential gender differences in overconfidence affecting compensation choices except following the first and last rounds. Subjects earned \$1 if their beliefs were correct in the randomly selected round used for payment.[4] Beliefs about ranking in the group were elicited before subjects chose a payment scheme to eliminate ex-post rationalization of their entry decisions.

At the end of the experiment, we collected data on additional personal characteristics in three different stages. We first elicited subjects' risk preferences using an incentivized senary lottery

---

[2] While a higher performance indirectly affects others' earnings by increasing performance thresholds of competitive tiers that others need to surpass in the next round, we do not suspect subjects to intentionally underperform because of concerns arising from this channel.

[3] Learning effects may lead to improved performance in the initial rounds. While these instances can occur because subjects of similar ability may be randomly grouped, we do not suspect collusions of this type since subjects cannot communicate and individual incentives prevent them from intentionally underperforming given the finitely repeated game.

[4] We did not suspect incentive compatibility concerns with such belief elicitation since incentives to exert effort are high enough in all payment schemes. Further, subjects may learn only the two highest scores of the other group members which eliminates concerns about the two lowest performers in the group possibly undercutting each other to guess their rank as being last. There are a total of 9 (out of 2440) instances in which subjects quit working and correctly guess their rank as being last. In case of ties in the actual ranks, we counted every answer that could be correct as correct. For example, if the performance in the group was 5, 5, 6, 6, then an answer of third and fourth was correct for a score of 5, and an answer of first and second was correct for a score of 6.

choice similar (scaled-down payments) to Eckel and Grossman (2008). We use this task to obtain a measure of risk preferences which could potentially affect compensation choices. Subjects selected one of six different lotteries (1-6) where higher numerical choices correspond with less risk-averse behavior. In the second stage, we elicited subjects' beliefs about the task stereotype regarding gender differences in performance which was not incentivized. In the last stage, subjects completed a demographic survey. A comprehensive summary of the sequence of tasks used to study the dynamics of competitive sorting behavior can be found in Table A1 in the Appendix.

During the experiment we relied on a facial expression analysis software called AFFDEX by Affectiva (McDuff et al., 2016; Stöckli et al., 2018) to assess subjects' affective responses while they elected one of three compensation schemes and during the feedback stage. This software captures slight changes in facial muscle movement and can measure the extent to which a subject experiences each of eight emotional states (joy, sadness, anger, disgust, contempt, fear, surprise and confusion). It is completely noninvasive and operates through any high-resolution webcam.[5]

3   Results

This section presents the performance of subjects in the real effort task and tournament entry decisions while accounting for different factors such as risk aversion and confidence. We describe the sample and basic performance first. We then analyze tournament entry choices absent any feedback and the dynamics of tournament entry choices following the provision of exogenously determined feedback. We finish by discussing potential explanations behind the gender gap in efficient sorting.

3.1   Sample description

A total of 244 Texas A&M undergraduate students participated in the experiment during Fall 2023 and Spring 2024 semesters in 26 gender-balanced sessions where 2 or 3 groups of 4 subjects partic-

---

[5]Once a subject's face is recognized by the software, it registers key facial landmarks that are used to detect emotion-related features. This information is then fed to a computer algorithm that uses a database of facial expressions of over 5.3 million faces from across the globe to determine the subject's emotional state (Zijderveld, 2017). AFFDEX then generates an index for each emotional state that reflects the degree to which the subject's expression aligns with a particular emotion. The index for each emotional state is unidirectional and ranges between 0 and 100, where 0 indicates no expression of a particular emotion (e.g., joy) whatsoever, and 100 is the highest level of confidence in a specific expression of emotion. AFFDEX has been previously validated as an accurate tool for measuring emotions (González-Rodríguez, Díaz-Fernández, & Gómez, 2020; Kulke, Feyerabend, & Schacht, 2020).

ipated in each session.[6] Table 1 presents means of observable characteristics by treatment. Given the downsides of balance testing (Ho, Imai, King, & Stuart, 2007), we compare the composition of the *Conditional-Disclosure* (N=124) and the *Unconditional-Disclosure* (N=120) treatments using Standardized Mean Differences (SMD) (Imbens & Rubin, 2015; Imbens & Wooldridge, 2009). According to Cochran and Rubin (1973)'s rule of thumb, the absolute value of SMD should be less than 0.25. As shown in Table 1, there are no meaningful differences between the *Conditional-Disclosure* and the *Unconditional-Disclosure* conditions in observable characteristics.

Table 1: SUMMARY STATISTICS AND NORMALIZED DIFFERENCES BETWEEN TREATMENTS

	Treatment		SMD
	Cond.-Disclosure	*Uncond.-Disclosure*	
Age	19.81	20.22	0.247
	(1.32)	(1.89)	
Male	0.50	0.50	0.000
	(0.50)	(0.50)	
Risk	3.45	3.38	-0.042
	(1.66)	(1.57)	
African American	0.09	0.08	-0.005
	(0.29)	(0.28)	
Asian/Pacific Islander	0.22	0.21	-0.009
	(0.41)	(0.41)	
Caucasian/White	0.49	0.45	-0.042
	(0.50)	(0.50)	
Hispanic	0.18	0.23	0.048
	(0.38)	(0.42)	
Native American/Indigenous	0.01	0	-0.008
	(0.09)	(0)	
Other Race	0.01	0.03	0.017
	(0.13)	(0.18)	
N	124	120	224

*Note*: Standard deviations reported in parentheses.

3.2   Performance in the Real Effort Task

In this section we examine if there are gender differences in performance in any of the 12 rounds which were comprised of 1 round of the piece-rate task, 1 tournament round and 10 rounds featuring choices over different compensation schemes that applied to future performance, in that order. In the piece-rate, the average number of problems solved is 4.52 and 4.89 for women and men, respectively; this difference is marginally significant based on a rank sum test (rank sum test,

---

[6]Gender was never mentioned to participants though they could observe the (balanced) session composition since they were seated in the same waiting room before entering the lab. We relied on the meta-analysis of gender differences in willingness to compete by Markowsky and Beblo (2022) for sample size estimation based on power analysis which can be found in the AEA RCT Registry (AEARCTR-0011911).

$p = 0.088$; two-sided t-test, $p = 0.147$). There is no gender difference under tournament incentives where women on average solved 5.37 problems and men solved 5.48 (rank sum test, $p = 0.599$; two-sided t-test, $p = 0.664$). Figure 1 shows the average score of women and men over the 12 rounds separated by treatment.[7] We observe that both women and men improve their performance over time at a similar pace in both conditions. Particularly, there are meaningful improvements in performance in early rounds, resulting from learning effects with performance gradually leveling over time.

Figure 1: AVERAGE NUMBER OF CORRECTLY SOLVED PROBLEMS
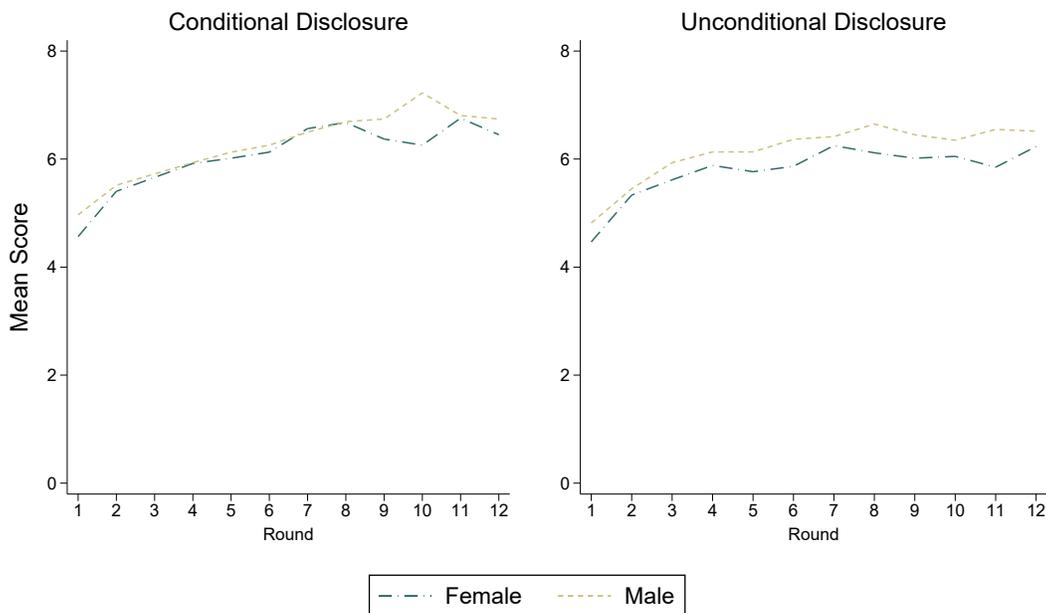
The similar performances of women and men should result in there being no gender difference in the probability of winning the Round 2 tournament. Of the 61 Round 2 tournaments, 26 were won by women and 35 were won by men. Conditioning only on gender, the probability of winning the tournament is 21.3 percent for a woman and 28.7 percent for a man; in a sample of 122 women and 122 men this difference is not significant ($p = 0.263$). Following the Round 2 tournament, we cannot reject the null hypothesis that the performances of women and men are comparable in the rounds that featured endogenously chosen payment schemes in either treatment (rank-sum tests, $p > 0.10$; two-sided t-tests, $p > 0.10$) with the exception of Round 10 in the *Conditional-Disclosure* condition where women on average solved 6.26 problems and men solved 7.23 (rank sum test,

---

[7]Even though feedback treatments were introduced at the conclusion of Round 3, there are no gender differences in performance in the first three rounds in either treatment.

$p = 0.040$; two-sided t-test, $p = 0.054$).[8] We further show that there are no gender differences in performance after splitting scores by quartile (see Table A2 in the Appendix). Taken together, the lack of a difference in performance by gender, for either treatment, removes one obvious potential reason for gender differences in compensation choices.

3.3 Gender Differences in Tiered Tournament Entry Before Feedback (Round 3 Choice)

The literature on gender differences in competitive behavior documents a robust finding that women are less likely to enter tournaments than men (see Markowsky and Beblo (2022) for a comprehensive summary). To test whether this basic pattern holds in our setting, we focus on compensation choices in Round 3 which features the first round where participants have a choice to enter one of the two tiered tournaments or not, aligning well with the environment used in previous work. In the majority of the studies from Markowsky and Beblo (2022)'s meta-analysis, participants did not have information about their relative performance, and in our study such feedback arrives only after Round 3. The available compensation choices of *Tier I*, *Tier II* or *piece-rate* can be ordered by decreasing competitiveness, with the higher tiered-tournament (lower number) being most competitive and the *piece-rate* scheme being least competitive. Table 2 shows the frequency count of each compensation scheme by gender. Despite the similar performances of women and men in the two initial rounds, their payment scheme choices are remarkably different. Though both women and men opt for the *Tier II* tournament at similar rates, we observe that the choices of women can be characterized as *Skipping-the-Top* tier, and of men as *Skipping-the-Bottom* non-competitive tier. These gender differences in tiered-tournament entry are both substantial and significant (Fisher's exact test, $p = 0.002$). This pattern of results replicates previous findings that gender differences in competitive choices are robust with the inclusion of a less competitive tournament.

Table 2: DISTRIBUTION OF COMPENSATION CHOICES BEFORE FEEDBACK (ROUND 3)

	Female	Male
Tier I	30	49
Tier II	52	55
Piece-rate	40	18
Total	122	122

---

[8]After adjusting for multiple comparisons, this difference is no longer significant under conventional levels (List, Shaikh, & Xu, 2019; Romano & Wolf, 2005).

While there are no gender differences in performance between women and men, other factors such as risk preferences or confidence as measured by one's guessed rank in the Round 2 tournament may influence compensation choices. Given this, we use an ordered probit model to test whether the gender differences in tiered-tournament entry remain after controlling for other potentially relevant factors. Table 3 shows that gender differences in competitive choices persist after adding these controls, along with the addition of other control variables for performance in the Round 2 tournament, and learning improvement from the first two rounds. As expected, both confidence and the actual performance in the Round 2 tournament are positively correlated with choosing more competitive payment schemes. The results remain practically unchanged when using performance from the first round instead. Improvement in performance between the first and second round (Improve (R2)), measured as the subject's performance from Round 2 subtracted by the performance in the first round, had no significant effect.[9] Column 2 includes a measure of risk aversion in a task similar to the one used by Eckel and Grossman (2008). We find this measure of risk aversion is significantly correlated with choosing more competitive payment schemes in Round 3. While women in our data are more risk-averse than men as most of the literature suggests, the magnitude of the effect for a female selecting less competitive compensation schemes slightly drops and its significance persists.

Table 3: ORDERED PROBIT OF CHOICES BEFORE FEEDBACK (ROUND 3)

	(1)	(2)
Female	-.297**	-.268**
	(.115)	(.118)
Guessed rank (R2)	-.493***	-.494***
	(.120)	(.117)
Improve (R2)	-.046	-.048
	(.047)	(.048)
Performance (R2)	.071*	.072*
	(.043)	(.043)
Risk control	No	Yes
Observations	244	244
Log-likelihood	-232.1	-231.5
χ^2	48.92	61.92

Notes: Dependent variable: round 3 choice of payment scheme (0-Piece rate, 1-Tier II, 2-Tier I). Standard errors clustered by experimental session and reported in parentheses.
* $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

---

[9]Subjects with a marked improvement in performance could be more likely to choose more competitive payment schemes than those whose performances are not positively affected by competitive settings. We do not find this pattern in our data.

In estimating marginal effects, we find that a female has a 0.083 lower probability of choosing the *Tier I* tournament than a male after controlling for performance, confidence and risk preferences. For a female to be as likely to choose the *Tier I* tournament as an average male, her rank belief about her relative performance would need to decrease by more than half a rank which represents a meaningful increase in confidence. An additional correct answer in Round 2 would increase the probability of entering the *Tier I* tournament by 0.023. A female would have to improve her score by a roughly 45 percentile increase from the median to be as likely to enter the *Tier I* tournament as a male. Hence, the gender differences in skipping the *Tier I* tournament are both significant and substantial.

Before examining feedback treatment effects on improving the accuracy of beliefs on relative performance and sorting efficiency into payoff-maximizing compensation schemes, we test whether two patterns are observed in our data based on previous findings: i) while both women and men are commonly overconfident about their relative performance, men tend to be more overconfident than women, and ii) men enter competitions at higher rates than is optimal and women commonly underenter competitions thereby foregoing higher earnings. We first confirm previous findings that, conditional on performance, women are less confident about their relative ranking than men and that subjects with a higher tournament performance believe they have higher relative performance (see Table A3). To investigate over (under) confidence, we create a variable based on actual and guessed rank as *actual rank − guessed rank*.[10] Similarly, to investigate over and underentry in choosing one of different compensation schemes, we encode the two types of inefficient sorting by creating a variable that equals 1 if a *Tier II* performer opts for the *Tier I* tournament, equals 0 if a *Tier II* performer opts for *Tier II*, and equals -1 if a *Tier II* performer opts for the *piece-rate* payment scheme.[11] Hence, a value of 0 indicates a payoff-maximizing choice, positive values indicate overentry and negative values indicate underentry. In the analyses, we use OLS regressions without a constant to calculate means for each gender. We omit the constant to test the efficiency of beliefs and entry choices for each gender, and to test for gender differences. Table 4 provides estimates of these effects.

Regarding overconfidence, we find partial evidence consistent with previous findings. While both women and men are commonly overconfident about their beliefs on past relative performance,

---

[10]Possible values for this variable range between -3 and +3. A value of 0 indicates accurate beliefs, positive values indicate overconfidence and negative values indicate underconfidence.

[11]This variable ranges between -2 and 0 for a *Tier I* performer, and between 0 and 2 for a *piece-rate* performance level.

in our data only men are overconfident ($p < 0.01$). In introducing a less competitive tournament in compensation choices, we find that both women and men overenter; that is, on average both opt for more competitive schemes than is optimal. Importantly, however, we find a statistically significant difference in overentry between women and men, with men overentering significantly more.

Table 4: PAYOFF-MAXIMIZING BELIEFS AND ENTRY CHOICES

	Over (Under) Confidence (Round 2)	Over (Under) Entry (Round 3)
Female	.082	.257
	(.075)	(.127)
	[.284]	[.053]
Male	.475	.495
	(.076)	(.124)
	[.000]	[.000]
H0: Female=Male	10.35	4.01
	[.004]	[.056]
Risk control	No	Yes
Observations	244	244

Notes: Standard errors clustered by experimental session and reported in parentheses. p-values in brackets.

3.4 Gender Differences in Tiered Tournament Entry Following Feedback (Rounds 4-12)

The literature on gender differences in competitive choices suggests that providing participants with feedback on relative performance eliminates or reduces the gender gap in competitiveness (see Markowsky and Beblo (2022), and the references therein). In implementing such an intervention, we randomly assigned groups to one of the two feedback conditions in a between-subject design. In the *Conditional-Disclosure* condition, although subjects received feedback on relative performance when electing to compete in one of the tiers, the opportunity cost of competing at the other tier was not disclosed.[12] In the *Unconditional-Disclosure* condition, subjects learned the opportunity cost of alternative payment schemes regardless of their compensation choices. Figure 2 presents the frequency of different compensation choices by treatment and gender. In the *Conditional-Disclosure* condition, gender differences in compensation choices appear to persist through the end of the experiment in all three compensation schemes. In the *Unconditional-Disclosure* condition, on the other hand, gender differences in compensation choices seem to get smaller over time in

---

[12]Subjects could always infer the opportunity cost of choosing the *piece-rate* since compensation was a linear function of the number of correctly solved problems in this scheme. When a subject opted for the *piece-rate* scheme, she/he did not learn the opportunity cost of having entered neither competitive tier.

Figure 2: Distribution of Compensation Choices Following Feedback (Rounds 4-12)

(a) *Conditional-Disclosure*
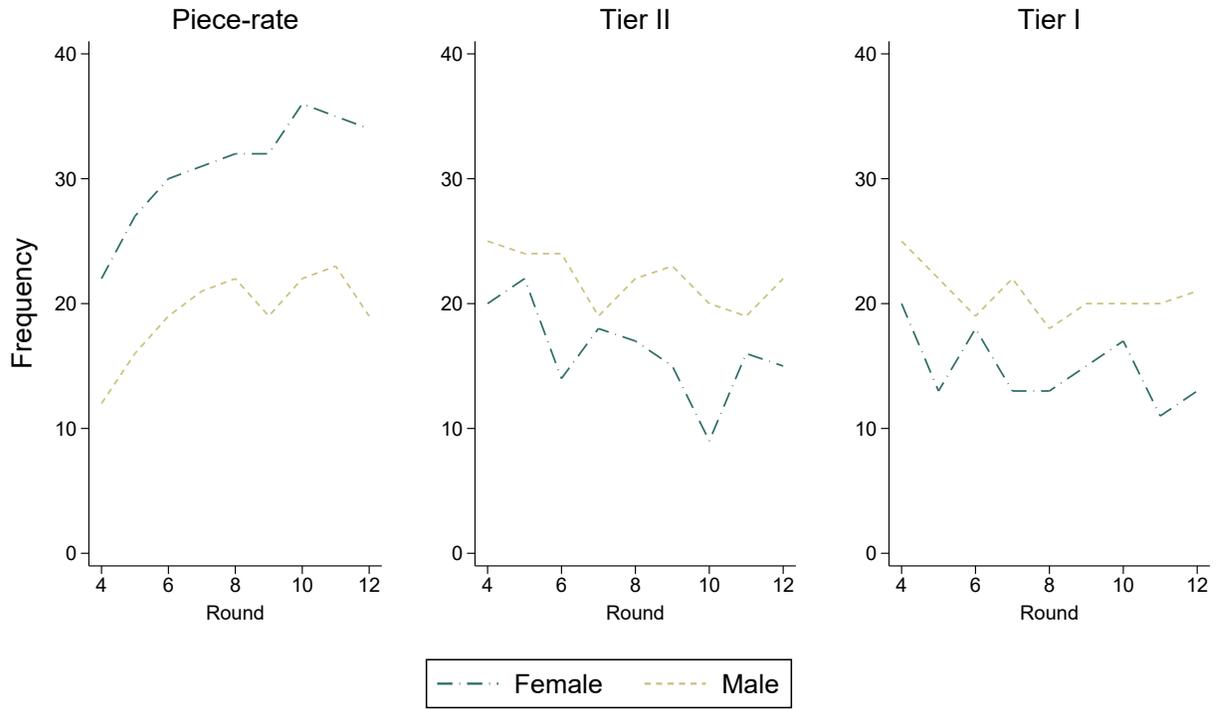
(b) *Unconditional-Disclosure*
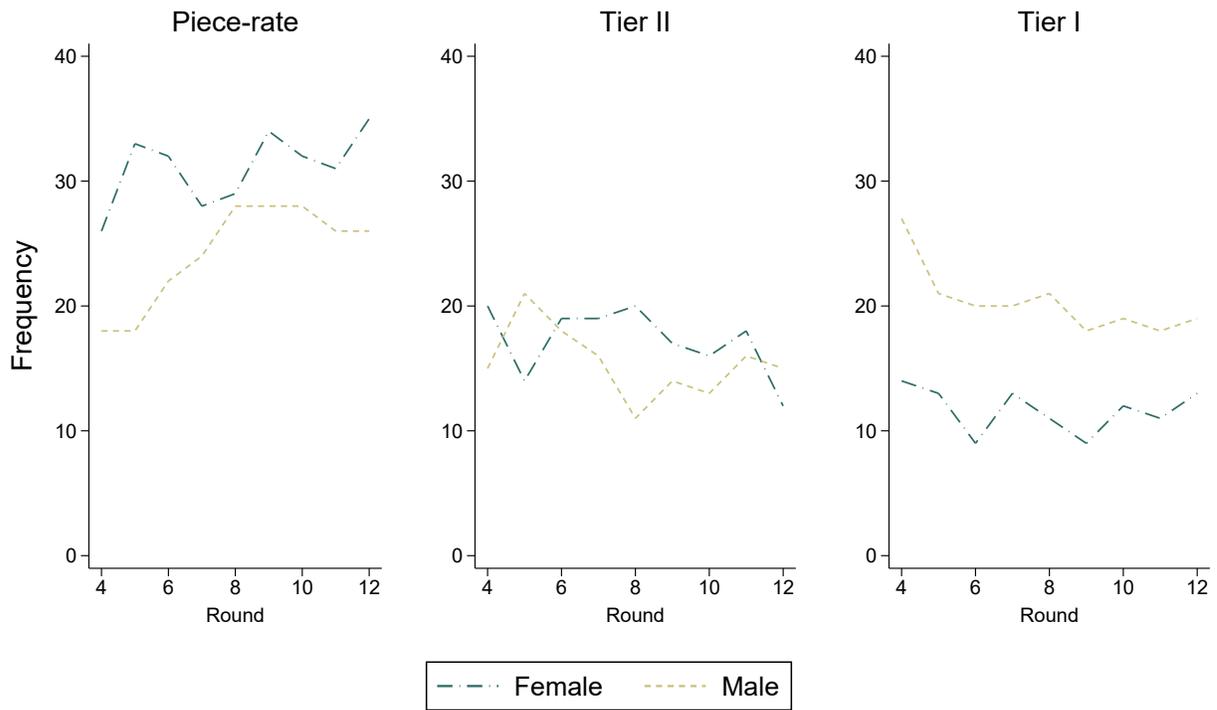
selecting the *piece-rate* scheme and the *Tier I* tournament, and it appears there is no difference in choosing the *Tier II* tournament over time. In testing whether women and men opt for different compensation choices, Fisher's exact tests largely confirm the observed patterns in the two conditions. In the *Conditional-Disclosure* condition, men opt for more competitive payment schemes than women through the end ($p < 0.1$, except in rounds 4, 7 and 8). In the *Unconditional-Disclosure* condition, women and men opt for similar payment schemes after the first few rounds with feedback ($p > 0.1$, except in rounds 4, 5, 6 and 8).

Table 5: ORDERED PROBIT OF CHOICES FOLLOWING FEEDBACK (ROUNDS 4-12)

	(1)	(2)
Guessed rank (R-1)	-.772***	-.772***
	(.040)	(.041)
Female	-.275**	-.275*
	(.138)	(.145)
Unconditional-Disclosure	-.059	-.059
	(.137)	(.094)
Female x	.071	.071
Unconditional-Disclosure	(.205)	(.210)
Clustering	Individual	Session
Clusters	224	26
Risk control	Yes	Yes
Round effects	Yes	Yes
Observations	2196	2196
Log-likelihood	-1877.8	-1877.8
χ^2	463.83	1001.26

Notes: Dependent variable: round 4-12 choices of payment scheme (0-Piece rate, 1-Tier II, 2-Tier I). Standard errors reported in parentheses.
* $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

To test the robustness of these results we use an ordered probit model to examine whether gender differences in payment scheme selections remain only in the *Conditional-Disclosure* condition and after controlling for rank beliefs on relative performance from the last round, risk preferences and individual round effects.[13] Table 5 shows that gender differences in competitive choices are apparent in the *Conditional-Disclosure* condition. Importantly, the effect is similar in magnitude to Round 3 where subjects made compensation choices without any feedback ($p = 0.046$).[14] In the *Unconditional-Disclosure* condition, there is no gender gap in competitive choices ($-0.275 + 0.071 = -0.204$, $p = 0.192$).[15] Comparing the difference in gender differences in competitive choices across

---

[13]Recall, we elicited incentivized beliefs about rank performance at the conclusion of each round before subjects chose a payment scheme.

[14]Clustering at the session-level yields $p = 0.059$.

[15]Clustering at the session-level yields $p = 0.170$.

the two treatments, information treatments do not have a meaningful effect. As expected, we observe that rank beliefs on relative performance are significantly correlated with compensation choices. In estimating marginal effects, we find that a female has a 0.071 lower probability of choosing the *Tier I* tournament than a male in the *Conditional-Disclosure* condition ($p = 0.044$), and a meaningful difference does not attain in the *Unconditional-Disclosure* condition (-0.052, $p = 0.192$).

Figure 3: Over (Under) Confidence Following Feedback

(a) *Conditional-Disclosure*

(b) *Unconditional-Disclosure*
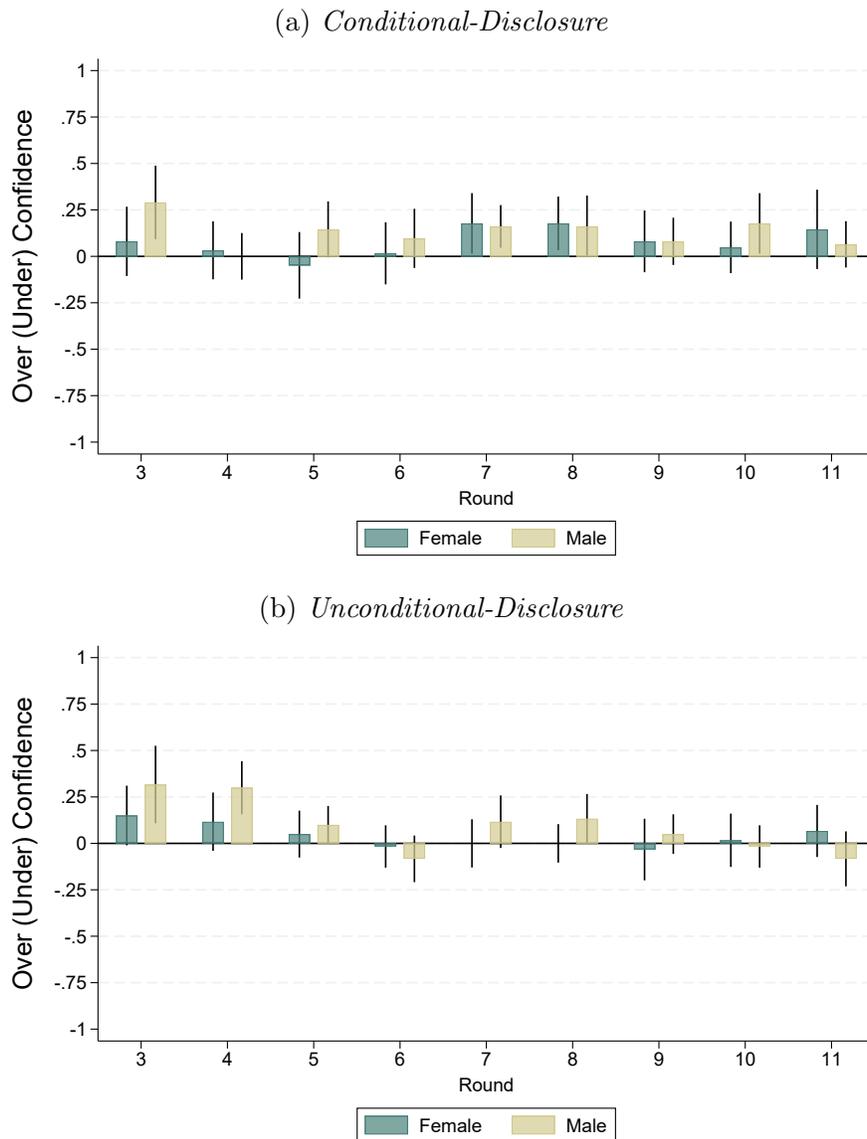
Before investigating the efficiency of competitive choices by gender, we first examine if the two treatments corrected the overconfidence of men observed in the Round 2 tournament (see Table 4).

We rely on this same variable introduced in section 3.3, namely the *actual rank − guessed rank*, for encoding over and underconfidence. Figure 3 presents average and 90% confidence intervals of this measure by experimental condition and gender.[16] We observe that following Round 4, in both conditions, the overconfidence of men appears to have been corrected, and both women and men seem to hold accurate beliefs about their past relative performance.

Figure 4: OVER (UNDER) ENTRY FOLLOWING FEEDBACK

(a) *Conditional-Disclosure*

(b) *Unconditional-Disclosure*
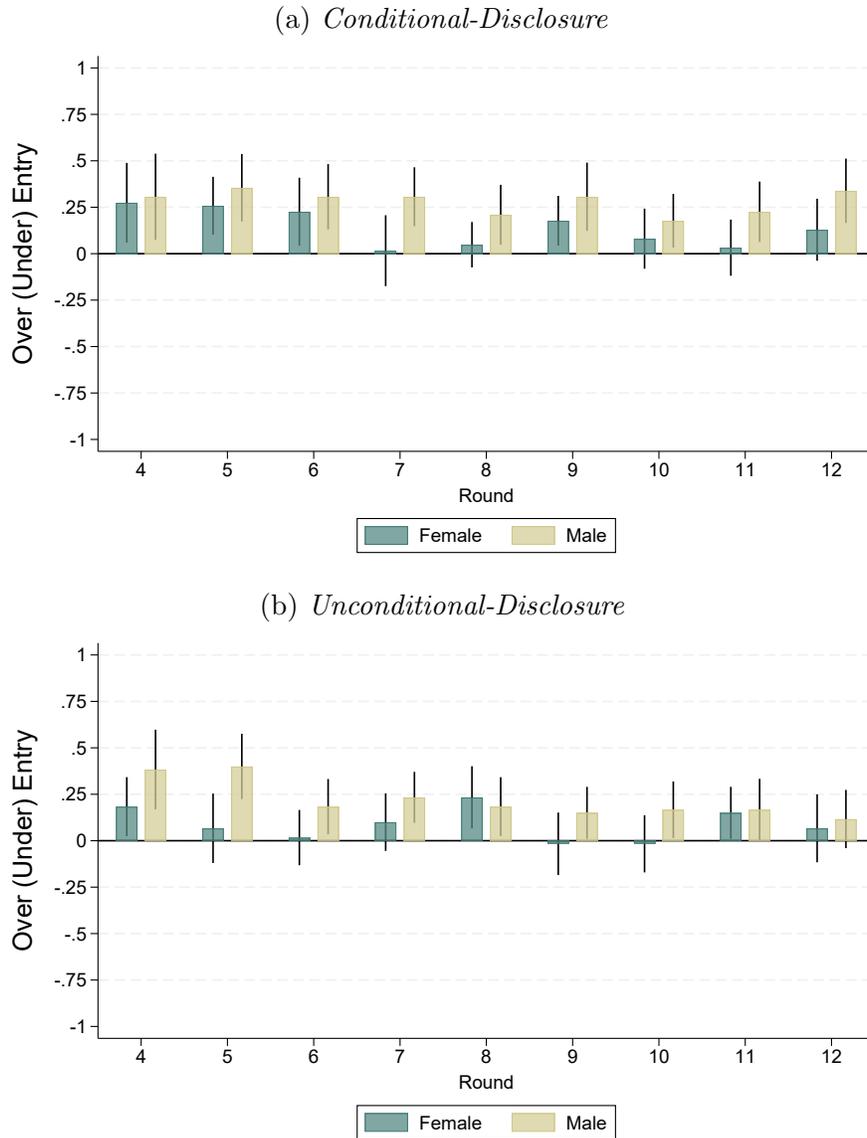
To investigate whether the gender differences in competitive choices are efficient, we examine whether subjects opted for payoff-maximizing compensation schemes given their relative perfor-

---

[16]Technically, beliefs following feedback were elicited after Round 3 since subjects indicated their beliefs before feedback was displayed.

mance in the group. We rely on the measure mentioned in section 3.3 for encoding over and underentry. Figure 4 presents average and 90% confidence intervals of this measure by experimental condition and gender. We observe that while women seem to opt for payoff-maximizing schemes in both conditions, men appear to efficiently sort only in the *Unconditional-Disclosure* condition.[17]

Thus far, the patterns observed in Figures 3 and 4 suggest that women hold accurate beliefs about past performance and opt for payoff-maximizing compensation schemes following feedback in both conditions. Whereas for men, these same patterns are observed only in the *Unconditional-Disclosure* condition. In the *Conditional-Disclosure* condition, men look as though they hold accurate beliefs about past performance but 'unrealistic expectations' about future performance as they continue to overenter in this condition through the end of the experiment. We pool observations following feedback to examine if these patterns are robust using OLS regressions without a constant and cluster errors at the individual level to account for repeated measures. Table 6 largely confirms these patterns. Although the overconfidence of men appears to have not been fully corrected in

Table 6: PAYOFF-MAXIMIZING BELIEFS AND ENTRY CHOICES FOLLOWING FEEDBACK

	Over (Under) Confidence (Rounds 4-11)			Over (Under) Entry (Rounds 4-12)		
	(1)	(2)	(3)	(1)	(2)	(3)
Male	.111	.144	.008	.125	.264	.252
	(.045)	(.076)	(.102)	(.070)	(.090)	(.103)
	[.014]	[.059]	[.937]	[.074]	[.004]	[.015]
Female	.079	.111	-.024	-.009	.131	.118
	(.056)	(.077)	(.099)	(.065)	(.088)	(.103)
	[.163]	[.151]	[.808]	[.896]	[.140]	[.250]
Unconditional-Disclosure	-.046	-.046	.229	-.077	-.077	-.052
	(.057)	(.057)	(.131)	(.064)	(.064)	(.128)
	[.417]	[.417]	[.080]	[.232]	[.232]	[.687]
Female x	-.007	-.007	-.007	.048	.048	.048
Unconditional-Disclosure	(.089)	(.089)	(.089)	(.088)	(.088)	(.088)
	[.934]	[.934]	[.934]	[.581]	[.581]	[.582]
Time trend		-.004	.012		-.017	-.016
		(.007)	(.011)		(.007)	(.009)
		[.583]	[.256]		[.011]	[.089]
Time trend x			-.032			-.003
Unconditional-Disclosure			(.014)			(.014)
			[.020]			[.817]
Risk control	No	No	No	Yes	Yes	Yes
Observations	1952	1952	1952	2196	2196	2196

Notes: Standard errors clustered at the individual level and reported in parentheses. p-values in brackets.

---

[17]We examined the possibility that there may be significant treatment or gender differences in performance fluctuation from initial levels and found no meaningful effects.

neither treatment (*Conditional-Disclosure*: 0.111, $p = 0.014$; *Unconditional-Disclosure*: $0.111 -$ $0.046 = 0.065$, $p = 0.067$), women and men have similar beliefs based on their actual and guessed rank performance following feedback in both conditions. In separate specifications, we include a time trend to account for beliefs and compensation choices potentially converging to optimum given the initially observed overconfidence and overentry (see Table 4).[18] We confirm that the lack of gender differences in rank performance beliefs relative to true rank are robust with the inclusion of a time trend. Specification (3) allows for the possibility that there are treatment differences in the rate at which beliefs converge to true ranks. Regression results show that an overall trend is not apparent in the *Conditional-Disclosure* condition (0.012, $p = 0.256$) although men become slightly overconfident with time in the *Conditional-Disclosure* condition ($p < 0.10$ after Round 6), but a negative and meaningful trend emerges in the *Unconditional-Disclosure* condition ($0.012 - 0.032 = -0.020$, $p = 0.022$). In this condition the overconfidence of both women and men is initially pronounced but is corrected after a few rounds with feedback ($p > 0.10$ after Round 6 for women, and $p > 0.10$ after Round 8 for men). In terms of electing payoff-maximizing payment schemes, we confirm that women's compensation choices are efficient while men overenter in the *Conditional-Disclosure* (0.125, $p = 0.074$) but not in the *Unconditional-Disclosure* condition ($0.125 - 0.077 = 0.048$, $p = 0.459$). When including a time trend, we observe a negative and significant trend in overentry as subjects choose payoff-maximizing payment schemes more frequently over time. The overall negative trend corrects the overentry of men only after Round 8 in the *Conditional-Disclosure* condition though Figure 4 shows men consistently overenter in this condition. Specification (3) shows that there are no differences in time trends between the two treatments as subjects' overentry decreases at similar rates in both conditions.

3.5 Explanations for the Gender Gap in Efficient Sorting

In section 3.4, we observed that women efficiently sort into payment schemes following feedback in both conditions. Whereas for men, this was the case only in the *Unconditional-Disclosure* condition. Table 7 presents average welfare loss from inefficient compensation choices due to both over and underentry. If subjects always pick the payoff-maximizing payment scheme, no welfare loss would have been incurred. We observe that men incur significantly more welfare loss than women in the *Conditional-Disclosure* condition, but this same result does not attain in the *Unconditional-*

---

[18]In results that follow, all gender differences are robust when controlling for individual round effects instead of relying on a time trend.

Disclosure condition where women and men incur similar welfare losses. This difference is primarily driven by men overentering in the *Conditional-Disclosure* condition as highlighted in Figure 4 and Table 6. Moreover, comparing the overall welfare loss by treatment, women are unaffected (two-sided t-test, $p = 0.171$), but men benefit significantly from unconditional disclosure (two-sided t-test, $p = 0.004$). The welfare loss difference in the gender difference across the two treatments is marginally insignificant ($p = 0.101$).

Table 7: Welfare Loss (\$) from Inefficient Compensation Choices

Round	Conditional-Disclosure			Unconditional-Disclosure		
	Female	Male	p-value	Female	Male	p-value
4	2.90	3.18	.645	1.89	2.57	.221
5	1.53	2.85	.028	1.82	1.65	.703
6	2.43	2.24	.760	1.49	2.07	.276
7	2.45	1.94	.397	1.36	1.52	.715
8	1.65	2.35	.263	1.74	1.67	.887
9	1.23	2.01	.103	1.78	1.58	.726
10	1.81	1.85	.956	1.73	1.37	.446
11	2.07	1.85	.695	1.56	1.71	.770
12	1.47	2.55	.051	1.98	1.82	.768
Overall	1.95	2.31	.061	1.71	1.77	.700

Notes: The table reports average welfare loss (in dollars) by treatment and gender. p-values are based on two-sided t-tests.

Our data convey a striking pattern in the *Conditional-Disclosure* condition where men continue to choose more competitive payment schemes than optimal through the end of the experiment. Although feedback largely corrects men's overconfidence regarding past relative rank performance, men hold unrealistic expectations about future performance. This pattern is consistent with evidence from a field study of managers working for a chain of food and beverage stores (Huffman, Raymond, & Shvets, 2022). The reason subjects may be motivated to hold unrealistic expectations might be because they gain utility indirectly from optimistic beliefs, or because confidence boosts motivation in improving performance (Bénabou & Tirole, 2016). In the *Unconditional-Disclosure* condition, subjects are provided with opportunity costs of alternative payment schemes regardless of their compensation choices and this intervention corrects men's overentry. In this condition, men (eventually) opt for payoff-maximizing compensation schemes after a few repeated rounds with feedback as they learn that downward mobility in competitive choices is efficient. An alternative explanation would be that medium and low ability men simply struggle accepting that downward mobility in competitive choices is optimal in the absence of repeatedly provided opportunity costs. Table 8 shows the probability of re-entering the same tournament following a loss in a previous

22

round. We observe that men, more than women, persist in re-entering tournaments where they previously experienced defeat in the *Conditional-Disclosure* condition, which is suggestive of *reality denial*. Despite receiving feedback revealing zero earnings following overentry, men continue to (over) compete more than women and incur significantly more welfare loss.

Table 8: PROBABILITY OF RE-ENTERING A TOURNAMENT FOLLOWING A PREVIOUS LOSS

(a) *Conditional-Disclosure*

	Tier I			Tier II		
Round	Female	Male	p-value	Female	Male	p-value
4	.15	.45	.082	.27	.56	.101
5	.23	.28	.708	.33	.30	.836
6	.38	.29	.435	.23	.47	.049
7	.30	.45	.204	.27	.34	.515
8	.24	.37	.260	.32	.33	.939
9	.29	.39	.391	.29	.44	.187
10	.32	.39	.555	.14	.39	.010
11	.22	.40	.102	.26	.33	.439
12	.22	.39	.118	.27	.45	.066
Overall	.27	.38	.005	.26	.40	.000

Notes: The table reports the probability of re-entering the same tournament after a loss in a previous round. p-values are based on two-sided t-tests.

(b) *Unconditional-Disclosure*

	Tier I			Tier II		
Round	Female	Male	p-value	Female	Male	p-value
4	.20	.54	.108	.18	.24	.683
5	.37	.32	.724	.16	.44	.031
6	.22	.36	.345	.32	.35	.793
7	.45	.43	.910	.36	.35	.929
8	.38	.45	.621	.40	.20	.070
9	.32	.41	.519	.32	.28	.670
10	.45	.44	.904	.33	.30	.740
11	.32	.42	.437	.37	.35	.832
12	.39	.49	.488	.27	.32	.648
Overall	.36	.43	.144	.32	.31	.966

Notes: The table reports the probability of re-entering the same tournament after a loss in a previous round. p-values are based on two-sided t-tests.

To further illuminate the mechanisms behind the observed overentry of men in the *Conditional-Disclosure* condition, the next section looks at facial expressions of subjects to assess their affective

responses while choosing one of three compensation schemes.[19]

3.5.1 Affective Evidence from Facial Expressions

To measure affective reactions for positive and negative emotions, we used a facial expression analysis software called AFFDEX by Affectiva (McDuff et al., 2016; Stöckli et al., 2018). The AFFDEX software operates at 30 Hz (i.e., it generates 30 observations per second). We constructed our AFFDEX measures as an average across these observations while subjects elected one of three compensation schemes. Figure 5 reports average and 90% confidence intervals of three different AFFDEX measures during the decision-making stage of subjects' competitive choices by treatment and gender: Valence, Anger and Sadness.[20] Valence tracks a subject's general emotional state, taking positive values when facial expressions indicate positive emotions and negative values when facial expressions indicate negative emotions. We observe that in both treatments women on average exhibit positive emotions whereas men on average exhibit negative emotions. Importantly, it appears that gender differences in Valence are meaningful and patterns are similar in both feedback conditions.[21] In the *Conditional-Disclosure* condition, we observe meaningful gender differences in Sadness in rounds 4 through 7 with men showing more sustained levels of sadness while women appear to regulate it (two-sided t-tests, $p < 0.1$);[22] these differences are not meaningful after adjusting for multiple comparisons. In rounds 7 and 8, men express meaningfully higher levels of Anger than women (two-sided t-tests, $p < 0.1$), which do not survive meaningful significance after adjusting for multiple comparisons. Mapping these gender differences in emotional expressions to the persistent overentry of men may indicate they struggled accepting that downward mobility presents a Pareto improvement as suggested by their increased levels of Anger halfway through the game. In the *Unconditional-Disclosure* condition, there appear to be no meaningful gender differences in expressed levels of Anger and an increasing trend in Sadness. This trend looks

---

[19]We further examine facial expressions during the feedback stage of outcomes in Appendix section A.1. Even if the monetary loss of different conditions are known, informational interventions entail other unobserved costs and benefits, making it difficult to assess welfare effects (Allcott & Greenstone, 2012; Loewenstein, 2000).

[20]Facial expression data were collected from 226 subjects. Similar figures showing round-by-round gender differences for the other six emotions can be found in the Appendix (see Figures A1 and A2).

[21]Meaningful differences are apparent in rounds 4, 5, 6, 7 and 12 in the *Conditional-Disclosure* condition (two-sided t-tests, $p < 0.1$); after adjusting for multiple comparisons, only the Round 5 difference is significant. Meaningful differences are apparent in rounds 7, 8, 9, 11 and 12 in the *Unconditional-Disclosure* condition (two-sided t-tests, $p < 0.1$); after adjusting for multiple comparisons, only the Round 9 difference is significant.

[22]Patterns in the overall emotional state, i.e. valence, suggest a positive trend for women. Women may be generating positive affect to a greater extent than men in order to down-regulate their negative responses, which is a pattern that has found empirical support (McRae, Ochsner, Mauss, Gabrieli, & Gross, 2008; Tamres, Janicki, & Helgeson, 2002; Thoits, 1991).

Figure 5: OVERALL EMOTIONAL STATE, ANGER AND SADNESS BY TREATMENT AND GENDER
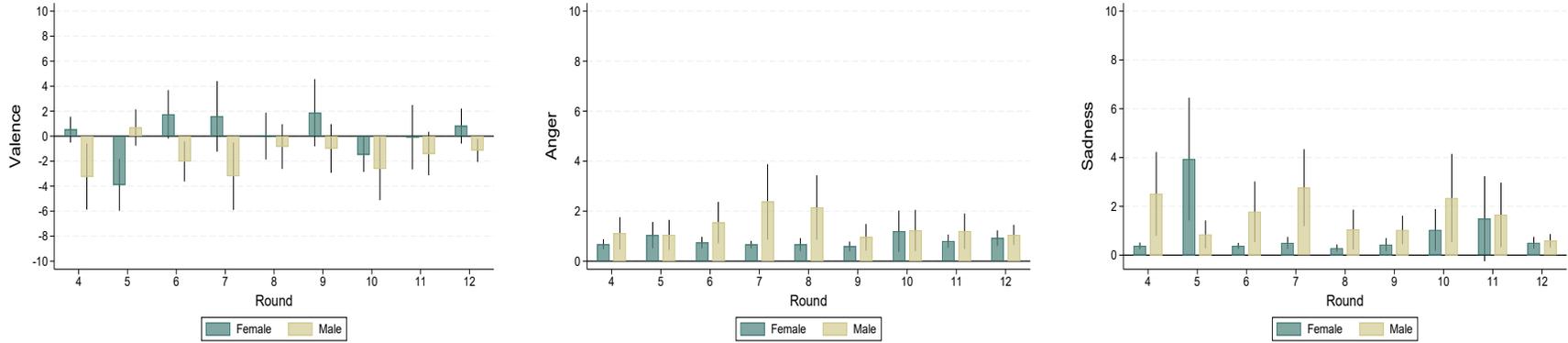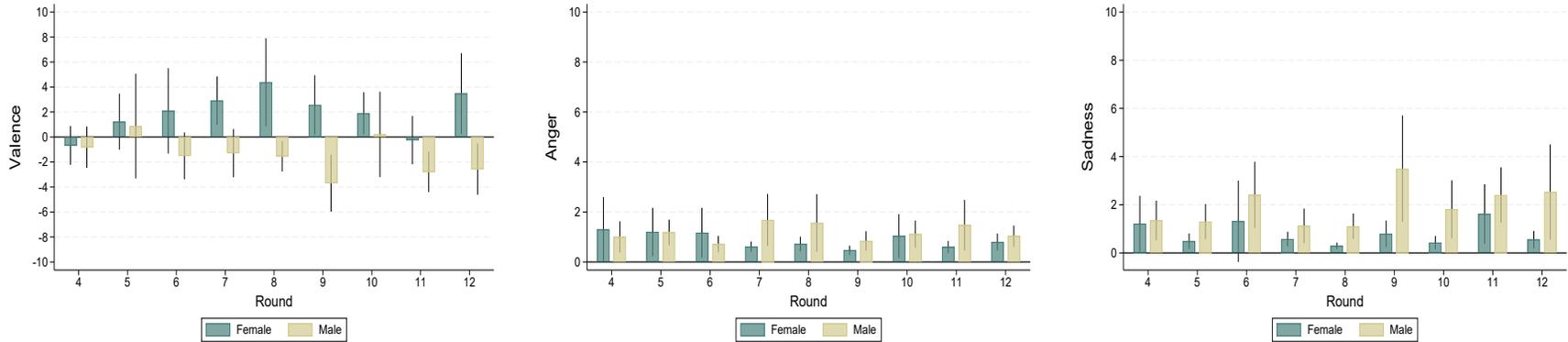
(a) *Conditional-Disclosure*

(b) *Unconditional-Disclosure*

Notes: The AFFDEX measure of Valence ranges from -100 to +100 and captures facial expressions of overall positive or negative emotions, with zero indicating neutral emotions. The AFFDEX measures of individual emotions are each on a scale from 0 to 100 and capture the level of confidence in expression of the indicated emotion. The device operates at 30 Hz, meaning it generates 30 observations per second for each subject. Each measure of a particular emotion is averaged for a given subject during the decision-making stage of competitive choices.

largely driven by men, while they also appear to be sadder than women.[23] Mapping these patterns with the observed decrease in overentry over time of men in the *Unconditional-Disclosure* condition is suggestive of acceptance that downward mobility is payoff-maximizing.[24]

Table 9 provides OLS regressions examining gender differences in the two treatments based on the AFFDEX measures of valence and eight individual emotions. We confirm that women express more positive emotions than men in both conditions (*Conditional-Disclosure*, 1.439, $p = 0.069$; *Unconditional-Disclosure*, $1.439 + 1.743 = 3.182$, $p = 0.002$). We further confirm that on average the expressed Anger levels of men are meaningfully higher than of women in the *Conditional-Disclosure* condition (0.840, $p = 0.051$), and this difference does not attain in the *Unconditional-Disclosure* condition ($0.840 - 0.554 = 0.286$, $p = 0.458$). A lack of a gender difference in expressed Anger levels in the *Unconditional-Disclosure* condition, but not in the *Conditional-Disclosure*, is indicative of male denial in the latter treatment which is already observed based on the persistent overentry of men. Opposite treatment patterns are observed for another negative emotion, namely Sadness. Men exhibit sadder facial expressions than women in the *Unconditional-Disclosure* condition ($0.430 + 0.632 = 1.062$, $p = 0.010$), but not under *Conditional-Disclosure* (0.430, $p = 0.359$). Similarly, when mapping these patterns of expressed emotions with gender differences in efficient sorting, men may experience sadder emotions under unconditional disclosure as a form of acceptance that downward mobility is economically efficient. Gender differences in the expression of other emotions do not appear to be meaningful under conventional levels except that men seem to be more confused than women in both conditions. One possible explanation behind this pattern is that men may have felt more conflicted about their compensation choices.

4   Conclusion

In this paper, we add tiered tournaments to the seminal study of Niederle and Vesterlund (2007) to investigate competitive sorting in a repeated game. The competitive tiers that underlie efficient sorting are chosen so that subjects who are below the median performance in the group should opt for the *piece-rate* scheme, and subjects in the top two quartiles appropriately sort in either the *Tier I* or *Tier II* tournament. Consistent with previous literature, we find men opt for more competitive schemes than women in this new environment before feedback interventions are introduced even

---

[23]Meaningful differences are apparent in rounds 5, 8, 9 and 10 (two-sided t-tests, $p < 0.1$); these differences are not meaningful after adjusting for multiple comparisons.

[24]For instance, in the Kübler-Ross model, sadness is experienced during the last stages of grief as a form of acceptance (Kübler-Ross, 1969).

Table 9: EMOTIONS DURING COMPENSATION CHOICES FOLLOWING FEEDBACK (ROUNDS 4-12)

	Valence	Anger	Sadness	Confusion	Contempt	Disgust	Fear	Joy	Surprise
Male	-1.439*	0.840*	0.430	1.031*	-0.644	1.589	1.848	-0.360	2.601
	(0.786)	(0.429)	(0.467)	(0.565)	(0.419)	(1.395)	(1.314)	(0.412)	(1.636)
Unconditional-Disclosure	1.685*	0.087	-0.290	-0.195	-0.373	1.589	-0.375	1.014	-0.489
	(0.987)	(0.264)	(0.336)	(0.292)	(0.485)	(1.395)	(0.780)	(0.851)	(0.867)
Male x	-1.743	-0.554	0.632	0.807	0.078	1.589	-0.838	-0.637	-1.008
Unconditional-Disclosure	(1.287)	(0.576)	(0.619)	(0.739)	(0.543)	(1.395)	(1.672)	(0.977)	(2.146)
Constant	-0.767	1.026***	1.567***	0.904***	2.124***	0.993*	1.521**	0.814*	1.560**
	(0.832)	(0.365)	(0.525)	(0.282)	(0.601)	(0.534)	(0.679)	(0.436)	(0.772)
Round effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2134	2134	2134	2134	2134	2134	2134	2134	2134

Notes: Standard errors clustered at the individual level and reported in parentheses.
* $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

after controlling for a host of influencing factors such as past performance, risk preferences and confidence as measured by the guessed rank performance in the last round. Consequently, men overenter competitive tiers significantly more than women. While both women and men are commonly overconfident about their relative performance, in our data only men are overconfident based on their actual and guessed rank performance.

Following the initial compensation choice made in the absence of any feedback, groups were randomly assigned into one of the two information treatments for the remaining rounds: *Conditional-Disclosure* or *Unconditional-Disclosure*. In the *Conditional-Disclosure* condition, although subjects received feedback on their earnings they did not learn the opportunity cost(s) of having selected (an) alternative payment scheme(s).[25] In the *Unconditional-Disclosure* condition, opportunity costs of alternative payment schemes were fully disclosed independent of compensation choices. We find that women hold accurate beliefs about past performance and opt for payoff-maximizing compensation schemes in both conditions. Whereas for men, neither treatment fully corrects their overconfidence on relative performance although its magnitude is significantly reduced from initial levels. Importantly, we do not find meaningful differences between women and men based on their actual and guessed rank beliefs in either treatment. In terms of efficiently sorting in one of different payment schemes, we find men's compensation choices are payoff-maximizing after a few rounds with feedback in the *Unconditional-Disclosure* condition. In the *Conditional-Disclosure* condition, on the other hand, men persistently overenter competitive tiers through the end of the experiment. We consequently documented that men benefit significantly from unconditional disclosure after examining welfare loss in the two information treatments. We discussed two potential explanations behind this pattern. First, while men hold largely accurate beliefs about past relative performance after a few rounds with feedback, they may hold unrealistic expectations about future performance because they gain utility from optimistic beliefs, or because confidence boosts motivation in improving performance. Second, we documented that men, more than women, persist in re-entering tournaments where they previously experienced defeat in the *Conditional-Disclosure* condition which could be suggestive of reality denial. Evidence from facial expressions while subjects chose one of three compensation schemes similarly suggests men may struggle with acceptance that downward mobility in competitive choices is Pareto efficient in the absence of repeatedly provided

---

[25]The only exception is when a subject selected *Tier II* and received feedback revealing zero earnings. In these instances, it is inferred that having chosen *Tier I* would have also yielded zero earnings, and the opportunity cost of having chosen the *piece-rate* could always be calculated since earnings are a linear function of the number of correctly solved problems in this scheme.

opportunity costs.

It remains an open question whether similar sorting patterns would be observed in labor markets in the field. Sorting is not only relevant between but also within firms. Firms can offer more transparent disclosures about promotions within and across different jobs to prod the right people toward the right jobs. Such disclosures could include performance expectations or productivity targets that can be objectively measured. Our results suggest transparency may especially benefit men in sorting into jobs where likelihood of advancement or promotion in a given career path is more likely.

References

Allcott, H., & Greenstone, M. (2012). Is there an energy efficiency gap? *Journal of Economic perspectives*, *26*(1), 3–28.

Alnamlah, M., & Gravert, C. A. (2020). She could not agree more: The role of failure attribution in shaping the gender gap in competition persistence.

Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, *95*(4), 1438–1443.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, *30*(3), 141–164.

Berlin, N., & Dargnies, M.-P. (2016). Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, *130*, 320–336.

Brandts, J., Groenert, V., & Rott, C. (2015). The impact of advice on women's and men's selection into competition. *Management Science*, *61*(5), 1018–1035.

Buser, T., & Yuan, H. (2019). Do women give up competing more easily? evidence from the lab and the dutch math olympiad. *American Economic Journal: Applied Economics*, *11*(3), 225–252.

Cason, T. N., Masters, W. A., & Sheremeta, R. M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, *94*(9-10), 604–611.

Chen, D. L., Schonger, M., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 417–446.

Cueva, C., Iturbe-Ormaetxe, I., Ponti, G., & Tomás, J. (2019). Boys will still be boys: Gender differences in trading activity are not due to differences in (over) confidence. *Journal of Economic Behavior & Organization*, *160*, 100–120.

Dargnies, M.-P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science*, *58*(11), 1982–2000.

Dasgupta, U., Mani, S., Sharma, S., & Singhal, S. (2019). Can gender differences in distributional preferences explain gender gaps in competition? *Journal of Economic Psychology*, *70*, 1–11.

Dohmen, T., & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American economic review*, *101*(2), 556–590.

Donovan, K., Lu, W. J., & Schoellman, T. (2023). Labor market dynamics and development. *The Quarterly Journal of Economics*, *138*(4), 2287–2325.

Dreber, A., Von Essen, E., & Ranehill, E. (2014). Gender and competition in adolescence: task matters. *Experimental Economics*, *17*, 154–172.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, *68*(1), 1–17.

Eeckhout, J. (2018). Sorting in the labor market. *Annual Review of Economics*, *10*, 1–29.

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

Elsby, M. W., & Gottfries, A. (2022). Firm dynamics, on-the-job search, and labor market fluctuations. *The Review of Economic Studies*, *89*(3), 1370–1419.

Ertac, S., & Szentes, B. (2011). *The effect of information on gender differences in competitiveness: Experimental evidence* (Tech. Rep.). Working paper.

Ewers, M. (2012). Information and competition entry.

Flory, J. A., Leibbrandt, A., & List, J. A. (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, *82*(1), 122–155.

Fluchtmann, J., Glenny, A., Harmon, N. A., & Maibom, J. (2024, May). The gender application gap: Do men and women apply for the same jobs? *American Economic Journal: Economic Policy*, *16*(2), 182-219. Retrieved from https://www.aeaweb.org/articles?id=10.1257/pol.20210607 doi: 10.1257/pol.20210607

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

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, *77*(5), 1637–1664.

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of economic literature*, *55*(1), 96–135.

Golman, R., & Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the

presence and absence of information. *Decision*, *5*(3), 143.

González-Rodríguez, M. R., Díaz-Fernández, M. C., & Gómez, C. P. (2020). Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics*, *51*, 101404.

Goulas, S., & Megalokonomou, R. (2021). Knowing who you actually are: The effect of feedback on short-and longer-term outcomes. *Journal of Economic Behavior & Organization*, *183*, 589–615.

Große, N. D., & Riener, G. (2010). *Explaining gender differences in competitiveness: Gender-task stereotypes* (Tech. Rep.). Jena Economic Research Papers.

Hahn, J. K., Hyatt, H. R., & Janicki, H. P. (2021). Job ladders and growth in earnings, hours, and wages. *European Economic Review*, *133*, 103654.

Healy, A., & Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal*, *121*(555), 1192–1204.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, *15*(3), 199–236.

Huffman, D., Raymond, C., & Shvets, J. (2022). Persistent overconfidence and biased memory: Evidence from managers. *American Economic Review*, *112*(10), 3141–3175.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, *47*(1), 5–86.

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

Kessel, D., Mollerstrom, J., & Van Veldhuizen, R. (2021). Can simple advice eliminate the gender gap in willingness to compete? *European Economic Review*, *138*, 103777.

Kübler-Ross, E. (1969). *On death and dying: What the dying have to teach doctors, nurses, clergy, and their own families.* Macmillan Publishing Co.

Kuhn, P., & Villeval, M. C. (2011). Do women prefer a co-operative work environment?

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A comparison of the affectiva imotions facial expression analysis software with emg for identifying facial expressions of emotion. *Frontiers in psychology*, *11*, 492813.

List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*, 773–793.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American economic review*, *90*(2), 426–432.

Markowsky, E., & Beblo, M. (2022). When do we observe a gender gap in competition entry? A meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, *198*, 139–163.

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. e. (2016). Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit. In *Proceedings of the 2016 chi conference extended abstracts on human factors in computing systems* (pp. 3723–3726).

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J., & Gross, J. J. (2008). Gender differences in emotion regulation: An fmri study of cognitive reappraisal. *Group processes & intergroup relations*, *11*(2), 143–162.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*, *68*(11), 7793–7817.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Romano, J. P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*(4), 1237–1282.

Shastry, G. K., Shurchkov, O., & Xia, L. L. (2020). Luck or skill: How women and men react to noisy feedback. *Journal of Behavioral and Experimental Economics*, *88*, 101592.

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, *10*(5), 1189–1213.

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with affdex and facet: A validation study. *Behavior research methods*, *50*, 1446–1460.

Tamres, L. K., Janicki, D., & Helgeson, V. S. (2002). Sex differences in coping behavior: A meta-analytic review and an examination of relative coping. *Personality and social psychology review*, *6*(1), 2–30.

Thoits, P. A. (1991). Gender differences in coping with emotional distress. In *The social context of coping* (pp. 107–138). Springer.

Topel, R. H., & Ward, M. P. (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics*, *107*(2), 439–479.

Tungodden, J., & Willén, A. (2023). When parents decide: Gender differences in competitiveness. *Journal of Political Economy*, *131*(3), 751–801.

Wasserman, M. (2023). Gender differences in politician persistence. *Review of Economics and Statistics*, *105*(2), 275–291.

Wozniak, D., Harbaugh, W. T., & Mayr, U. (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*, *32*(1), 161–198.

Zijderveld, G. (2017). The world's largest emotion database: 5.3 million faces and counting. *Affectiva. Luettavissa: https://blog. affectiva. com/the-worlds-largest-emotion-database-5.3-million-faces-and-counting Luettu*, *28*, 2018.

Appendix

Table A1: DESIGN OF THE EXPERIMENT

Round 1	Round 2		Round 3				Round ...	Round 12			End
Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step ...	Step 40	Step 41	Step 42	Step 43
Working time	Working time	Beliefs	Scheme choice	Working time	Beliefs	Feedback	...	Scheme choice	Working time	Feedback	Question-naires
3 mins	3 mins			3 mins		30 secs	...		3 mins	30 secs	
Piece-rate	Tourna-ment	Rank bel-ief in Step 2	Tier I vs. Tier II vs. Piece-rate	Tier I, Tier II or Piece-rate	Rank bel-ief in Step 5	Condi-tional or Uncondi-tional Disclosure	...	Tier I vs. Tier II vs. Piece-rate	Tier I, Tier II or Piece-rate	Condi-tional or Uncondi-tional Disclosure	Incentivi-zed Risky Choice Task Ste-reotype Elicitation Demogra-phics

Table A2: Descriptive Analysis of Performance

Round	Quartile	Conditional-Disclosure			Unconditional-Disclosure		
		Female	Male	p-value	Female	Male	p-value
1	$P(P_0 \leq Score \leq P_{25})$	.306	.210	.218	.317	.250	.418
	$P(P_{25} < Score \leq P_{50})$	.403	.403	1.000	.367	.300	.439
	$P(P_{50} < Score \leq P_{75})$	.097	.210	.081	.200	.317	.144
	$P(P_{75} < Score \leq P_{100})$	.194	.177	.817	.117	.133	.783
2	$P(P_0 \leq Score \leq P_{25})$	.290	.274	.842	.400	.317	.341
	$P(P_{25} < Score \leq P_{50})$	.290	.242	.542	.117	.167	.432
	$P(P_{50} < Score \leq P_{75})$	.242	.323	.318	.350	.367	.849
	$P(P_{75} < Score \leq P_{100})$	.177	.161	.811	.133	.155	.793
3	$P(P_0 \leq Score \leq P_{25})$	.290	.355	.442	.283	.317	.690
	$P(P_{25} < Score \leq P_{50})$	.161	.193	.638	.333	.283	.553
	$P(P_{50} < Score \leq P_{75})$	.403	.194	.011	.233	.200	.658
	$P(P_{75} < Score \leq P_{100})$	.145	.258	.117	.150	.200	.471
4	$P(P_0 \leq Score \leq P_{25})$	.306	.274	.692	.283	.317	.690
	$P(P_{25} < Score \leq P_{50})$	.306	.355	.567	.300	.250	.540
	$P(P_{50} < Score \leq P_{75})$	.177	.145	.625	.300	.217	.297
	$P(P_{75} < Score \leq P_{100})$	.210	.226	.828	.117	.217	.142
5	$P(P_0 \leq Score \leq P_{25})$	.290	.242	.542	.283	.250	.680
	$P(P_{25} < Score \leq P_{50})$	.258	.323	.429	.333	.283	.553
	$P(P_{50} < Score \leq P_{75})$	.290	.226	.412	.183	.217	.648
	$P(P_{75} < Score \leq P_{100})$	.161	.210	.488	.200	.250	.512
6	$P(P_0 \leq Score \leq P_{25})$	.242	.306	.421	.400	.350	.572
	$P(P_{25} < Score \leq P_{50})$	.387	.194	.018	.183	.217	.648
	$P(P_{50} < Score \leq P_{75})$	.210	.290	.300	.333	.217	.152
	$P(P_{75} < Score \leq P_{100})$	.161	.210	.488	.083	.217	.041
7	$P(P_0 \leq Score \leq P_{25})$	.290	.387	.255	.350	.317	.699
	$P(P_{25} < Score \leq P_{50})$	.258	.145	.117	.217	.167	.487
	$P(P_{50} < Score \leq P_{75})$	.258	.226	.675	.217	.367	.071
	$P(P_{75} < Score \leq P_{100})$	.194	.242	.514	.217	.150	.345
8	$P(P_0 \leq Score \leq P_{25})$	.371	.339	.707	.283	.333	.553
	$P(P_{25} < Score \leq P_{50})$	.177	.145	.625	.533	.300	.010
	$P(P_{50} < Score \leq P_{75})$	.226	.274	.534	.083	.133	.378
	$P(P_{75} < Score \leq P_{100})$	.226	.242	.832	.100	.233	.050
9	$P(P_0 \leq Score \leq P_{25})$	.403	.338	.457	.300	.250	.540
	$P(P_{25} < Score \leq P_{50})$	.306	.323	.847	.317	.333	.845
	$P(P_{50} < Score \leq P_{75})$	.145	.081	.256	.200	.183	.817
	$P(P_{75} < Score \leq P_{100})$	.145	.258	.117	.183	.233	.500
10	$P(P_0 \leq Score \leq P_{25})$	.452	.290	.063	.300	.233	.409
	$P(P_{25} < Score \leq P_{50})$	.226	.290	.412	.417	.483	.463
	$P(P_{50} < Score \leq P_{75})$	.129	.113	.783	.150	.083	.255
	$P(P_{75} < Score \leq P_{100})$	.194	.306	.147	.133	.200	.327
11	$P(P_0 \leq Score \leq P_{25})$	.323	.258	.429	.417	.317	.256
	$P(P_{25} < Score \leq P_{50})$	.306	.339	.701	.217	.167	.487
	$P(P_{50} < Score \leq P_{75})$	.129	.161	.610	.233	.300	.409
	$P(P_{75} < Score \leq P_{100})$	.242	.242	1.000	.133	.217	.230
12	$P(P_0 \leq Score \leq P_{25})$	.387	.339	.575	.367	.317	.564
	$P(P_{25} < Score \leq P_{50})$	.177	.129	.455	.350	.317	.699
	$P(P_{50} < Score \leq P_{75})$	.242	.274	.681	.100	.133	.570
	$P(P_{75} < Score \leq P_{100})$	.194	.258	.390	.183	.233	.500

Notes: The table reports the proportion of scores within each quartile by treatment, round and gender. p-values are based on two-proportion tests.

Table A3: ORDERED PROBIT OF GUESSED TOURNAMENT RANK (ROUND 2)

	(1)	(2)
Female	.740***	.731***
	(.149)	(.143)
Improve (R2)	-.037	-.037
	(.044)	(.044)
Performance (R2)	-.313***	-.313***
	(.040)	(.040)
Risk control	No	Yes
Observations	244	244
Log-likelihood	-262.4	-262.4
χ^2	85.73	85.88

Notes: Dependent variable: guessed rank in the round 2 tournament for guesses of ranks 1, 2, 3 and 4. Standard errors clustered by experimental session and reported in parentheses. * $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.
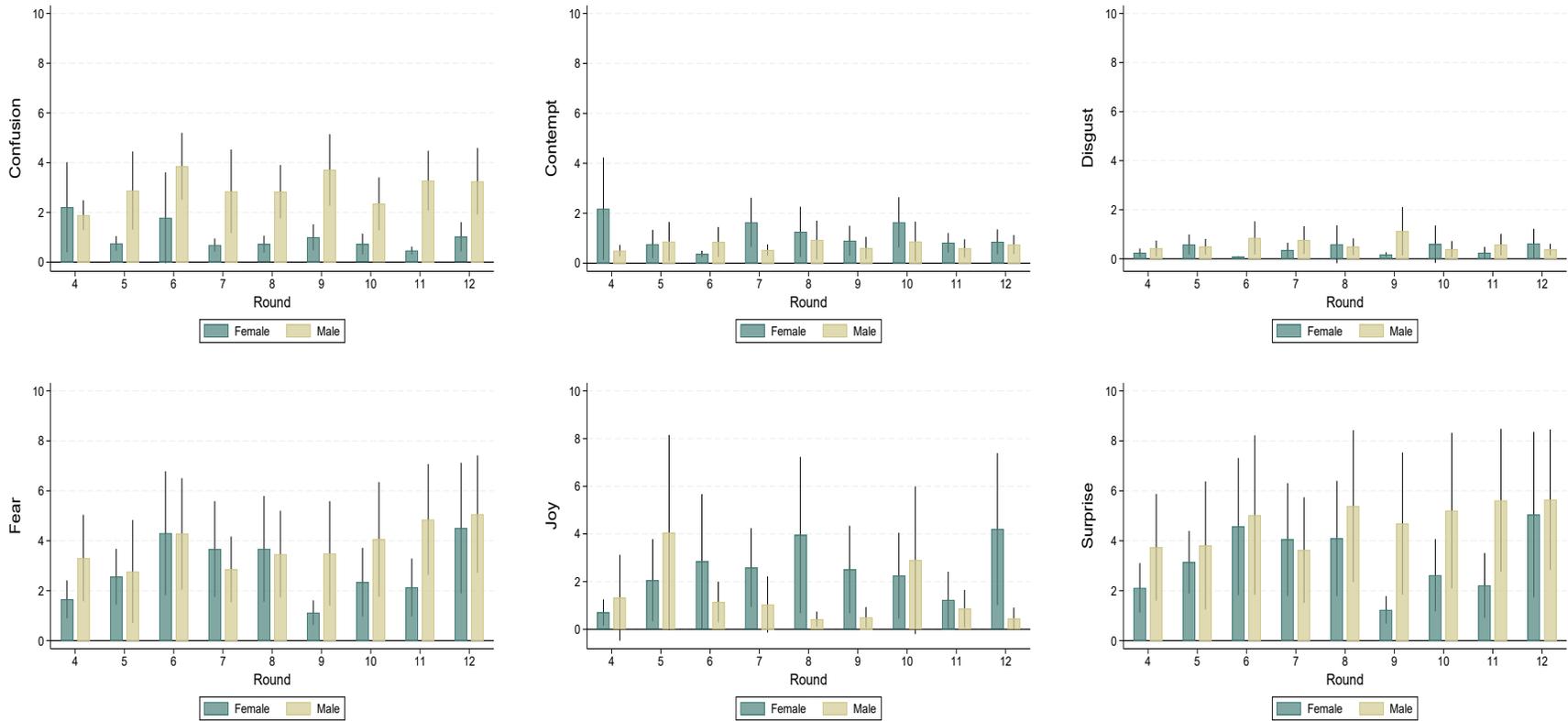
Figure A1: CONFUSION, CONTEMPT, DISGUST, FEAR, JOY AND SURPRISE (*Conditional-Disclosure*)
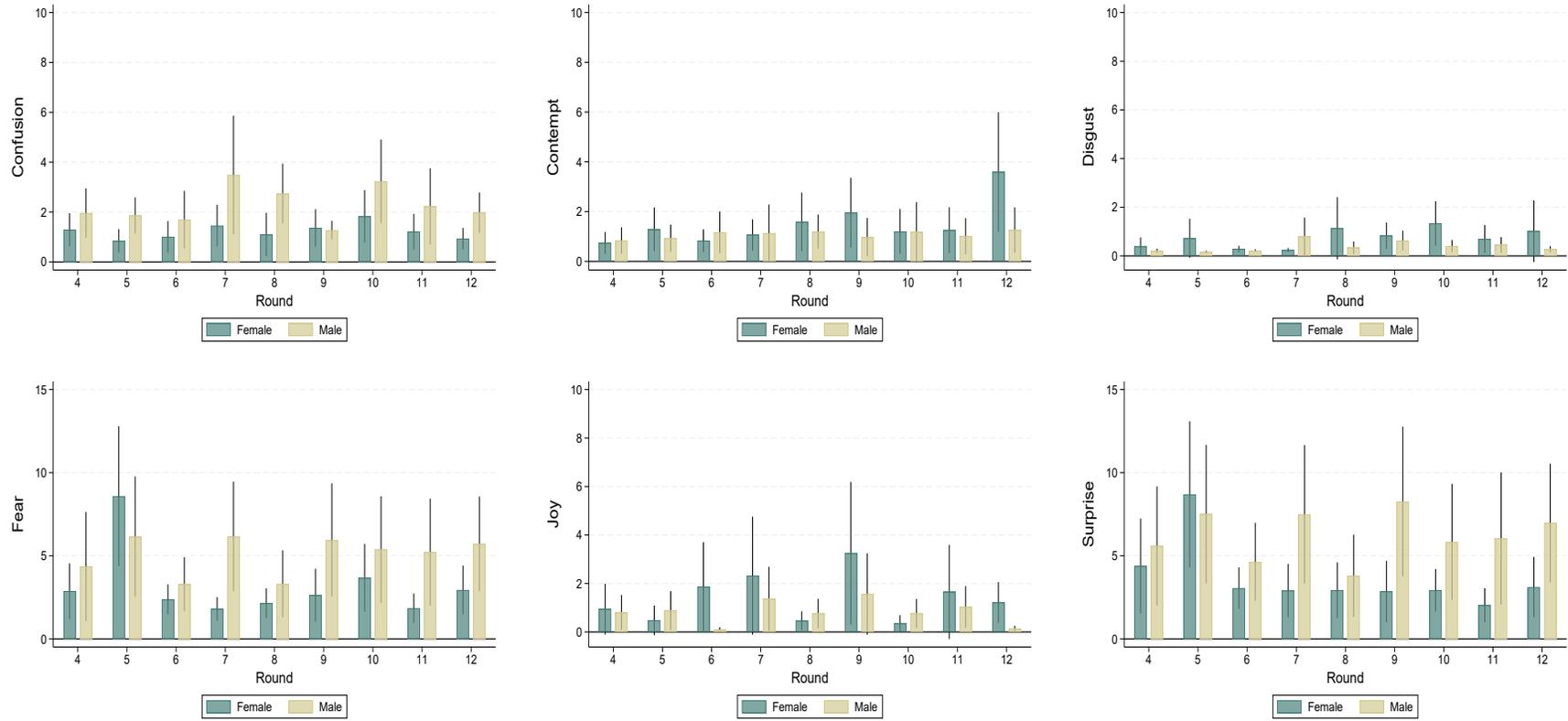
Notes: The AFFDEX measures of individual emotions are each on a scale from 0 to 100 and capture the level of confidence in expression of the indicated emotion. The device operates at 30 Hz, meaning it generates 30 observations per second for each subject. Each measure of a particular emotion is averaged for a given subject during the decision-making stage of competitive choices.

Figure A2: Confusion, Contempt, Disgust, Fear, Joy and Surprise (*Unconditional-Disclosure*)

Notes: The AFFDEX measures of individual emotions are each on a scale from 0 to 100 and capture the level of confidence in expression of the indicated emotion. The device operates at 30 Hz, meaning it generates 30 observations per second for each subject. Each measure of a particular emotion is averaged for a given subject during the decision-making stage of competitive choices.

A.1 Facial Expressions During the Feedback Stage

In this section, we examine facial expressions of subjects upon receiving feedback on their outcomes. Following Round 2, feedback was displayed for 30 seconds at the conclusion of each round. Figure A3 reports average and 90% confidence intervals of three different AFFDEX measures during the feedback stage by treatment and gender: Valence, Sadness and Disgust.[26] We observe that both women and men appear to exhibit neutral emotions under *Conditional-Disclosure*.[27] Under *Unconditional-Disclosure*, on the other hand, women appear to exhibit more positive emotions than men.[28] It is already worth noting that men incur significantly less monetary loss under *Unconditional-Disclosure* as observed in Table 7, while showing no difference in Valence between treatments. Although women do not significantly benefit monetarily from unconditional disclosure, their overall emotional state shows a marked improvement when opportunity costs of alternative payment schemes are disclosed independent of compensation choices. In terms of Sadness, we observe that women and men exhibit similar levels under *Conditional-Disclosure*,[29] whereas men appear sadder than women under *Unconditional-Disclosure* with differences especially pronounced in rounds 9, 11 and 12 (two-sided t-tests, $p < 0.1$); these differences do not survive significance after adjusting for multiple comparisons. Although women appear to show higher levels of Disgust than men on average under *Conditional-Disclosure*, there appear to be no round-by-round gender differences in this condition with the exception of Round 9 (two-sided t-test, $p = 0.029$). Meaningful gender differences in the expressed levels of Disgust are not apparent under *Unconditional-Disclosure*.

Table A4 provides OLS regressions examining gender differences in the two treatments based on the AFFDEX measures of valence and eight individual emotions. We confirm that women express more positive emotions than men only in the *Unconditional-Disclosure* condition ($1.047 + 1.716 = 2.763$, $p = 0.059$; *Conditional-Disclosure*, 1.047, $p = 0.247$).[30] As expected, men exhibit sadder facial expressions than women in the *Unconditional-Disclosure* condition ($0.340 + 0.594 = 0.934$, $p = 0.048$), but not in the *Conditional-Disclosure* (0.340, $p = 0.451$). We further confirm that on average the expressed Disgust levels of women are meaningfully higher than of men in the

---

[26]Similar figures showing round-by-round gender differences for the other six emotions are available upon request.

[27]No meaningful differences appear between women and men with the exception of Round 5 (two-sided t-test, $p = 0.07$).

[28]Meaningful differences are apparent in rounds 6, 9, 10 and 11 (two-sided t-tests, $p < 0.05$). None of the differences are statistically meaningful after adjusting for multiple comparisons.

[29]The only exception is Round 8 where men exhibit higher levels of Sadness than women (two-sided t-test, $p = 0.089$).
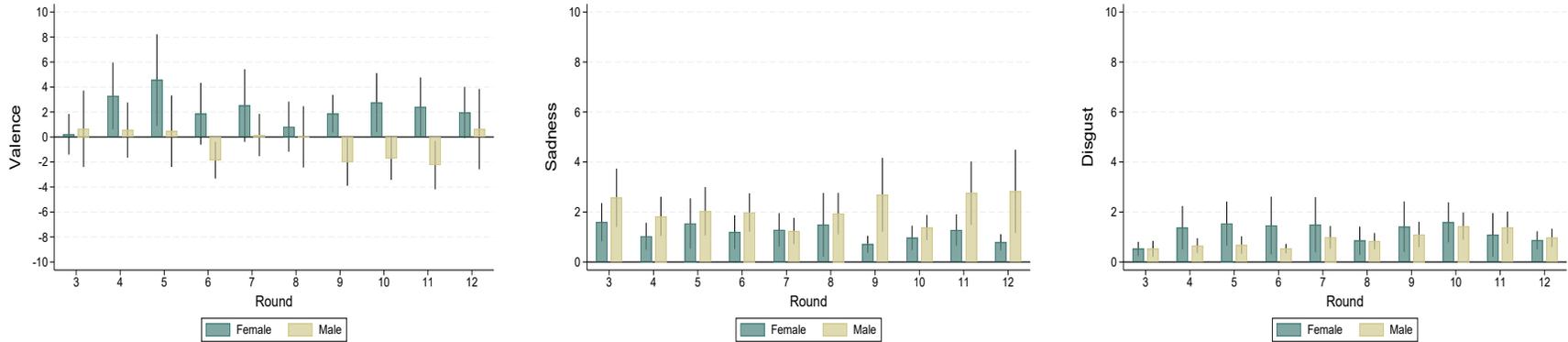
[30]We further examined possible treatment or gender differences in response to receiving negative (or positive) feedback after competing but found no meaningful effects.

Figure A3: Overall Emotional State, Sadness and Disgust by Treatment and Gender
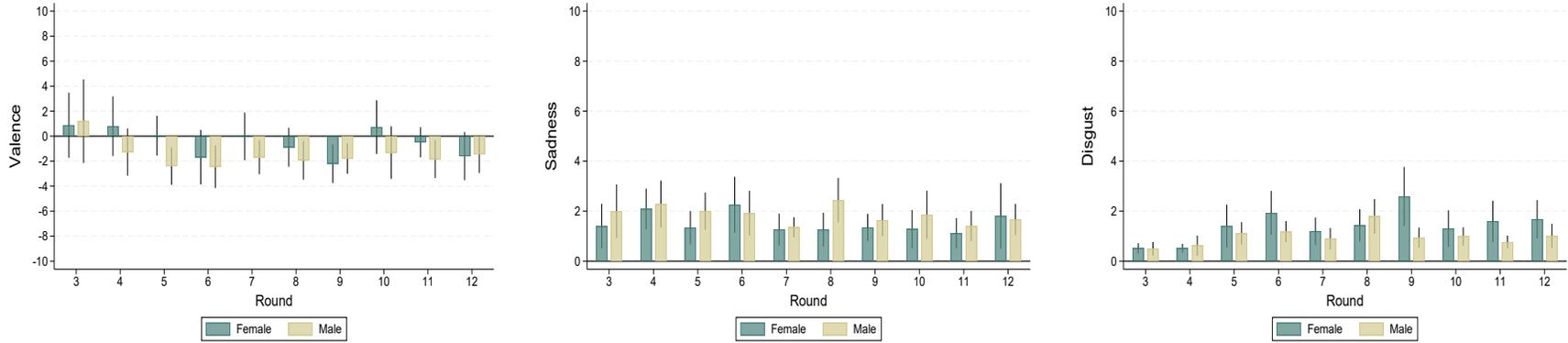
(a) *Conditional-Disclosure*

(b) *Unconditional-Disclosure*

Notes: The AFFDEX measure of Valence ranges from -100 to +100 and captures facial expressions of overall positive or negative emotions, with zero indicating neutral emotions. The AFFDEX measures of individual emotions are each on a scale from 0 to 100 and capture the level of confidence in expression of the indicated emotion. The device operates at 30 Hz, meaning it generates 30 observations per second for each subject. Each measure of a particular emotion is averaged for a given subject during the feedback stage where subjects learn about their outcomes.

Conditional-Disclosure condition (0.438, $p = 0.084$), and this difference does not attain in the main condition (0.438 − 0.130 = 0.308, $p = 0.452$).[31]

Involving our biometric estimates in computing welfare effects yields an identical recommendation to relying solely on minimizing monetary loss. If we make the heroic assumption that facial expressions upon learning one's outcomes are fully representative of utility, moving from conditional to unconditional disclosure presents a Pareto improvement. While men are unaffected, women show a marked improvement on their emotional state when opportunity costs of alternative payment schemes are disclosed independent of compensation choices.[32] Alternatively, if there is some degree of separability in utility being derived from earnings and experienced emotions, a similar policy recommendation would be made as both women and men incur less monetary loss under unconditional disclosure with men benefiting significantly more.

---

[31]Gender differences in the expression of other emotions do not appear to be meaningful under conventional levels upon receiving feedback except that men seem to be more confused than women in both conditions.

[32]Although women incur similar monetary loss in the two treatments, under unconditional disclosure earning gaps are evident. Compared to men, it is possible women may use positive emotions in the service of reappraising negative emotions to a greater degree to regulate their emotions (McRae et al., 2008).

Table A4: Emotions during the Feedback Stage (Rounds 3-12)

	Valence	Anger	Sadness	Confusion	Contempt	Disgust	Fear	Joy	Surprise
Male	-1.047	0.451	0.340	0.977**	-0.581	-0.438*	1.590	-0.305	2.278
	(0.902)	(0.289)	(0.450)	(0.460)	(0.416)	(0.252)	(1.321)	(0.569)	(1.503)
Unconditional-Disclosure	2.681**	0.088	-0.328	-0.163	-0.433	-0.201	-0.524	1.454	-0.194
	(1.300)	(0.261)	(0.439)	(0.357)	(0.444)	(0.441)	(0.793)	(1.129)	(0.924)
Male x	-1.716	-0.273	0.594	0.347	0.181	0.130	-0.163	-0.603	-0.617
Unconditional-Disclosure	(1.702)	(0.441)	(0.651)	(0.727)	(0.502)	(0.481)	(1.709)	(1.402)	(2.063)
Constant	0.370	1.002***	1.734***	1.273***	1.701***	0.804***	3.356***	2.821***	3.932***
	(0.992)	(0.190)	(0.439)	(0.265)	(0.369)	(0.216)	(0.813)	(0.803)	(0.898)
Round effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2180	2180	2180	2180	2180	2180	2180	2180	2180

Notes: Standard errors clustered at the individual level and reported in parentheses.
* $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

Experimental Materials

Figure A4: Conditional-Disclosure – Feedback Stage Example

Round 3 Results

Your Score: **0**

Time left on this page: **0:20**

The second-highest score of the other group members from the *previous round*:
Tier II Score: **0**

Selected Payment Scheme: **Tier II**
Earnings: **$0.00**

Figure A5: Unconditional-Disclosure – Feedback Stage Example

Round 3 Results

Your Score: **0**

Time left on this page: **0:20**

The two highest scores of the other group members from the *previous round*:
Tier I Score: **0**
Tier II Score: **0**

Selected Payment Scheme: **Tier II**
Earnings: **$0.00**

Instructions and Protocol

Welcome and thank you for participating in today's study.

You will receive compensation of $10.00 for your participation today. You may also receive additional payments depending on your decisions, the decisions of others, and luck. As a reminder, your participation is completely voluntary. You may choose to end your participation at any time. However, in order to receive your payment, you must complete the entire study. All information collected today will be confidential and will not be used for any reason other than this research.

Before we begin, please ensure your cell phone is turned off (or silenced) and all your belongings are placed near you on the floor. Please remain quiet and keep your eyes on your own screen for the duration of the study. All instructions will be given to you on your screen. If you have any questions, please raise your hand and a staff member will assist you.

Today you will be asked to solve sets of addition problems for 12 rounds without using a calculator. At the end of the experiment, one of these 12 rounds will be randomly chosen to determine your earnings; with each round being equally likely to be drawn.

The method we use to determine your earnings varies across rounds. Before each round, we will describe in detail how your payment is determined.

Each participant is randomly assigned to a 4-person group. Your group composition will stay the same for all of the 12 rounds.

Click *Next* when you are ready to proceed.

New Page

(Attention Check)

Which of these is not an animal? To indicate that you are paying attention, please choose 'Cow.'

→ Monkey, **Cow**, Lettuce, Whale

New Page

Round 1

In this round, you will be asked to solve sets of addition problems. You will be given 3 minutes to solve as many problems as you can. You cannot use a calculator; however, you are welcome to use the provided paper to compute your answers. Your answers are anonymous.

If this round is selected for payment, you will earn $0.50 for every problem you solve correctly. We

refer to this payment as the **piece-rate** payment.

(Working time)

Round 2

In this round, you will be asked to solve sets of addition problems. You will be given 3 minutes to solve as many problems as you can. You cannot use a calculator; however, you are welcome to use the provided paper to compute your answers. Your answers are anonymous.

If this round is selected for payment, you will earn $2.00 per problem answered correctly <u>**only**</u> if you are the **top performer** of your group and earn $0.00 if you are not the top performer. If at the end of the game there are any ties, a winner will be chosen randomly.

(Working time)

Round 2

In this round, how do you think your score ranked relative to the other participants in your group? If this round is selected for payment, you will earn $1 for guessing correctly.

→ First, Second, Third, Fourth

Rounds 3-12

In rounds 3-12, you will be asked to solve sets of addition problems. In each round, you will be given 3 minutes to solve as many problems as you can. You cannot use a calculator; however, you are welcome to use the provided paper to compute your answers. Your answers are anonymous.

As a reminder, groups were randomly formed at the beginning of the study and you are assigned to the same 4-person group for the entire session.

Before each round, you will select a payment scheme that has different performance requirements.

- **Tier I**: You will earn $2.00 per problem answered correctly if your score is higher than the score of the other 3 group members from the previous round and earn $0.00 if you do not outperform their scores.

46

- **Tier II**: You will earn $1.00 per problem answered correctly if your score is higher than the score of at least 2 other members of your group from the previous round and earn $0.00 if you do not outperform these scores.

- **Piece-rate**: You will earn $0.50 per problem answered correctly.

New Page

Comprehension Quiz

Reminder: groups were randomly formed at the beginning of the study and you are assigned to the same 4-person group for the entire session.

1. Earning money in Tier I requires that my score is higher than the score of _____ from the previous round.

   → 1 other group member, 2 other group members, **all 3 other group members**

2. Earning money in Tier II requires that my score is higher than the score of _____ from the previous round.

   → at least 1 other group member, **at least 2 other group members**, all 3 other group members

New Page

Round 3

In this round, you will be asked to solve sets of addition problems. You will be given 3 minutes to solve as many problems as you can. You cannot use a calculator; however, you are welcome to use the provided paper to compute your answers. Your answers are anonymous.

If this round is selected for payment, you will be paid according to the payment scheme you choose below:

- **Tier I**: You will earn $2.00 per problem answered correctly if your score is higher than the score of the other 3 group members from *Round 2* and earn $0.00 if you do not outperform their scores.

- **Tier II**: You will earn $1.00 per problem answered correctly if your score is higher than the score of at least 2 other members of your group from *Round 2* and earn $0.00 if you do not outperform these scores.

- **Piece-rate**: You will earn $0.50 per problem answered correctly.

Please pick a payment scheme:

→ Tier I, Tier II, Piece-rate

If you have any questions, please raise your hand.

New Page

(Working time)

New Page

Round 3

In this round, how do you think your score ranked relative to the other participants in your group? If this round is selected for payment, you will earn $1 for guessing correctly.

→ First, Second, Third, Fourth

New Page

Round 3 Results (*feedback*)

New Page

Round 4

In this round, you will be asked to solve sets of addition problems. You will be given 3 minutes to solve as many problems as you can. You cannot use a calculator; however, you are welcome to use the provided paper to compute your answers. Your answers are anonymous.

If this round is selected for payment, you will be paid according to the payment scheme you choose below:

- **Tier I**: You will earn $2.00 per problem answered correctly if your score is higher than the score of the other 3 group members from *Round 3* and earn $0.00 if you do not outperform their scores.

- **Tier II**: You will earn $1.00 per problem answered correctly if your score is higher than the score of at least 2 other members of your group from *Round 3* and earn $0.00 if you do not outperform these scores.

- **Piece-rate**: You will earn $0.50 per problem answered correctly.

Please pick a payment scheme:

If you have any questions, please raise your hand.

New Page

(Working time)

New Page

Round 4

In this round, how do you think your score ranked relative to the other participants in your group? If this round is selected for payment, you will earn \$1 for guessing correctly.

$\rightarrow$ First, Second, Third, Fourth

New Page

Round 4 Results (*feedback*)

.....

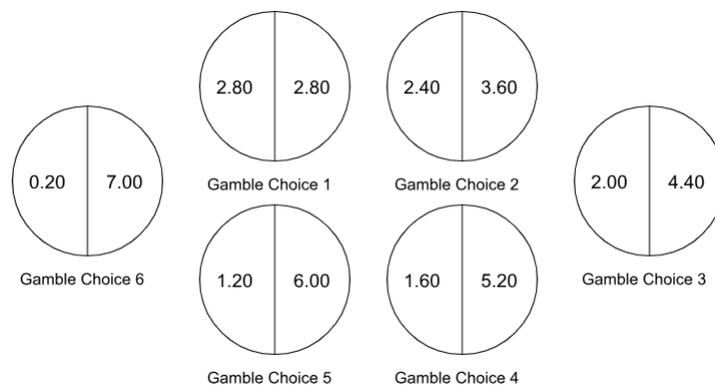(A similar sequence of pages follows until **Round 12**)

New Page

Gamble Task

You must choose one and only one of these gambles. For each gamble, two outcomes are equally likely. There are no right or wrong answers, you should select the gamble option that you like best. Please indicate your choice below.

You will be paid for this task at the end of the experiment.

0.20	7.00	Gamble Choice 6
2.80	2.80	Gamble Choice 1
2.40	3.60	Gamble Choice 2
2.00	4.40	Gamble Choice 3
1.20	6.00	Gamble Choice 5
1.60	5.20	Gamble Choice 4

$\rightarrow$ 1, 2, 3, 4, 5, 6

(*Task stereotype elicitation*)

Do you think men or women generally solve more problems in the addition task you did today?

→ Men, Women, No difference

Please answer the following questions.

1. Age

   →

2. Major

   →

3. Sex

   → Male, Female

4. Race

   → Asian/Pacific Islander, African American, Caucasian/White, Hispanic, Native American/Indigenous, Other

(*Summary of subject earnings*)