Relative feedback response in competitive environments[1]

Diego Aycinena
Department of Economics
Universidad del Rosario
diego.aycinena@urosario.edu.co

Rimvydas Baltaduonis
Department of Economics
Gettysburg College
rbaltadu@gettysburg.edu

Haylee Downey
Department of Psychology
Utah State University
hayleedowney@outlook.com

Kristian Fors
Department of Economics and Finance
Utah State University
krfors@gmail.com

Lucas Rentschler
Department of Economics and Finance
Utah State University
lucas.rentschler@usu.edu

[1] We thank the Department of Economics at Gettysburg College for generous financial support.

Abstract

Relative performance feedback is common in competitive environments. Such feedback, especially when it is negative, is more likely to affect beliefs in women Roberts and Nolen-Hoeksema (1989). We investigate the effect of relative performance feedback on subsequent performance in a winner-take-all environment. We find that feedback slightly reduces the performance of men. Women, on the other hand, respond to negative feedback by increasing their subsequent performance, even when doing so is unlikely to affect the likelihood of winning.

JEL Classifications: L0, C9.

Keywords: gender, feedback, competition.

1 Introduction

Relative performance is often used to determine outcomes in the workplace, with prominent examples including bonuses, promotions, and performance evaluations (Prendergast, 1999). It is often used to determine which employee will obtain some desirable prize. That is, there are pecuniary incentives to being at the top of the distribution.[1]

Since using relative performance feedback is so ubiquitous in competitive environments, it is important to understand the effect of the *content* of such feedback on subsequent performance. Such information is relevant to belief formation about relative ability (see e.g., Krähmer (2007)), and thus the likelihood of being at the top of the distribution. Thus, feedback could lead to complacency on the part of top-ranked employees, and could lead to the discouragement of low ranked employees. Alternatively, feedback could inspire increased performance, particularly at the top of the distribution, as high-ranked individuals vie for dominance. Performance is also likely to be affected by information on the relative ranking, since it leads to social comparison (Garcia et al., 2013), and a person's relative position may affect their self-image (Eil and Rao, 2011; Zimmermann, 2020) or their self-identity (Tesser and Campbell, 1980). This is particularly important since Ertac (2011) shows that negative relative performance feedback is over-weighted when it is relevant to self-evaluation.

The effects of feedback content may also differ by gender. Some research from psychology has shown that women are more likely to internalize negative feedback (see e.g., Roberts and Nolen-Hoeksema, 1989).[2] Indeed, in the context of competition Niederle and Vesterlund (2007) notes that "...women, more than men, may view a negative signal as indicative of their self-worth rather than simply their one-time performance on a task." While the effects of relative performance feedback may be more likely to impact women, the effect on their behavior within a competition is unclear.[3] Are women more likely to become discouraged and reduce their performance in response to negative feedback? If women are more likely to take negative feedback to heart, would it lead to an increase in performance?

We evaluate this question in a laboratory setting. Our experimental design is related to that of Niederle and Vesterlund (2007), with several important differences. Subjects in fixed groups engage in several rounds of a real-effort task which consists of solving addition problems for five minutes. The initial rounds involve a piece-rate payment scheme followed by a winner-take-all tournament. They then choose linear combinations of the two preceding schemes, thus providing measures of competitiveness. Beliefs regarding their relative ranks are elicited. Before participating in another tournament, subjects are told their rank in the initial tournament. Finally, risk preferences are elicited. Our primary interest is in how the performance of subjects responds to the

[1] While the use of such incentives is thought to inspire better performance (Prendergast, 1999), unintended consequences can emerge. Charness et al. (2014) finds that people are willing to sabotage competitors in an effort to improve their own place in the ranking in an environment where the ranking did not affect payment. Azmat and Iriberri (2016) finds that negative feedback can also reduce job satisfaction under a piece-rate payment scheme.

[2] Baldwin et al. (2003) shows that women respond emotionally to both positive and negative cues more strongly than men. Huang et al. (2018) finds evidence that the way in which feedback is framed can have varying levels of effectiveness across genders. They find that feedback framed in a cooperative sense is most effective for motivating women, and that competitively framed feedback is most effective for motivating men.

[3] Wozniak et al. (2014) shows that the more accurate beliefs induced by relative performance feedback eliminates the gender difference in entry into competitive environments. See Niederle (2016) for an excellent review of the literature on gender differences in preferences for competition.

content of the feedback they receive.

We find that, overall, feedback does not significantly affect subsequent performance. However, this is driven by opposing gender effects. Women increase their performance, while men tend to reduce their performance. Further, the content of feedback only significantly impacts the performance of women. Women increase their performance only when they receive negative feedback. That is, women who did not rank first in the previous tournament increase their subsequent performance. The magnitude of this performance increase is highest for bottom ranked women, which suggests that their response is not driven by monetary considerations.

2 Related literature

We are not the first to consider the effect of feedback on performance in competitive environments. Hannan et al. (2008) find that under a tournament incentive scheme, subjects perform better with no feedback, whereas under an individual incentive scheme, feedback increases the performance mean. In their experiment, subjects did not engage in a real-effort task, so that the received feedback is less likely to be taken as relevant for self-image. Eriksson et al. (2009) focus on the role of providing relative performance feedback within a single competition, and find that such feedback does not improve overall performance. They find a similar result under a piece-rate payment scheme.

Several other laboratory experiments consider the effect of relative performance feedback in non-competitive environments. Azmat and Iriberri (2016) consider the effect of feedback under both piece-rate and fixed payment schemes, and finds that feedback only increases the performance of men under a piece-rate scheme. Charness et al. (2014) finds that, in flat-wage environment, people will sabotage others to improve their own standing in the rankings. Also in an flat-wage environment Kuhnen and Tymula (2012) shows that providing relative performance feedback improves the accuracy of beliefs about future rank. They also find that those with a high rank tend to reduce output, while those with a low rank tend to increase output. Gill et al. (2019) report the results of an experiment which is, to the best of our knowledge, the only other analysis of the effect of the content of relative performance feedback on subsequent performance. They find a u-shaped effect, in which the low and high ranked increase their performance, while intermediate ranks lower performance. This experiment did not involve a competition; subjects received a fixed payment for their participation. Notably, they find no gender difference in the performance response to feedback.

There are also several field experiments which consider feedback in environments where rankings have no direct impact on outcomes. Barankay (2011) show that in a workplace crowd-sourcing experiment, employees who received information about their performance rank were less productive and less likely to continue working, while Barankay (2012) demonstrates that in a sales environment, removing rank feedback increased sales performance. Azmat and Iriberri (2010) find that providing information to students about the average class performance increases students' grades by 5%. Azmat et al. (2019), however, demonstrated that such grade transparency decreases performance.

3 Experimental design

In each session, twelve subjects are randomly and anonymously sorted into three groups of four. These groups are held fixed for the duration of the experiment. Video instructions, which consist of slides with an experimenter reading the text, are shown on each subject's computer.[4] Subjects are provided with headphones.

Subjects participate in a series of activities, and watch the relevant instructions immediately before each activity. They do not know what future activities involve beforehand, although they do know that several will involve a specific real-effort task. Subjects do not receive feedback about the outcome of any activity during the experiment, except in one instance, explained below. After the experiment has concluded, one of the activities is randomly chosen for payment via bingo cage. The outcome of this activity is displayed to subjects, so that they understand how their payments are determined.

The real effort task used in this experiment consists of summing five two-digit numbers. Specifically, five two-digit numbers are chosen at random and displayed on a subject's screen. Subjects are asked to enter the sum. Once an answer is entered, five new numbers are chosen. A counter on the screen indicates the number of correct answers the subject has submitted. After five minutes, the task concludes.[5]

The sequence of our design is similar to that of Niederle and Vesterlund (2007). First, subjects familiarize themselves with the software interface by participating in a practice round of the real-effort task. The results of this activity cannot affect a subject's earnings.

After completing the practice round, Activity 1 consists of the real-effort task under a piece-rate payment scheme in which they earn $GTQ5$ per correct answer.[6] In Activity 2, subjects again participate in the real-effort task. This time, the payment scheme is a winner-take-all tournament, in which subjects earn $GTQ20$ per correct answer if they have the highest number of correct answers in their group of four (ties are resolved in favor of the subject who submitted the answer fastest). If they do not have the highest number of correct answers they earn $GTQ0$.

Activity 3 again involves participating in the real-effort task. Before the task begins, however, each subject chooses a payment scheme that is a linear combination of the piece-rate scheme and the tournament scheme. As in Niederle and Vesterlund (2007), the performance of a subject is compared against their three group members in Activity 2. The weight a subject puts on the tournament scheme is a measure of their competitiveness. Notably, this measure is less granular than that of Niederle and Vesterlund (2007), which used a binary choice between the two payment schemes.[7]

In Activity 3, subjects first chose their payment scheme, and then participate in the real effort task. This gives them the opportunity to change their behavior in the real-effort task in response to the chosen payment scheme. To obtain a measure of competitiveness that does not allow such a response, Activity 4 involves subjects choosing a payment scheme (again a linear combination of

[4] See Appendix A for the text of the instructions.

[5] On average, the performance of men and women does not differ in this task. See Table A1 in Appendix B for summary statistics and the result of non-parametric tests for each of the instances in which subjects engage in this task during our experiment.

[6] Experimental payments were denominated in local currency units (GTQ). At the time of the experiment, exchange rate was $USD1 \approx GTQ7.81$.

[7] This measure of competitiveness has also been employed in Saccardo et al. (2017).

the piece-rate and tournament schemes) that will be applied to the (already determined) outcome of Activity 1.

Activity 5 elicits beliefs about how a subject's performed relative to their group members in Activity 1 and Activity 2. Specifically, subjects are asked to report their beliefs about their rank within their group for both of these activities, with one being the highest rank, and four being the lowest rank. In the event that this activity is chosen for payment, one of the two questions is randomly chosen.[8] A subject receives $GTQ60$ if their reported belief is correct, and receives $GTQ10$ otherwise. By comparing believed ranks with realized ranks, we obtain a measure of overconfidence.

Activity 6 once again involves the real-effort task with the tournament payment scheme. However, there is one crucial difference between it and Activity 2. In particular, before the subjects begin, they are told their realized rank in Activity 2. Our interest is in determining how the performance of subjects changes in the face of this information, and whether or not there is a gender difference in this performance response.

Activity 7 is a standard risk elicitation task similar to Holt and Laury (2002), but adapted to mimic closely the possible outcomes of the different payment schemes.[9] In the event that this activity is randomly chosen for payment, one of the ten lottery choices is selected at random suing a ten-sided die, and the corresponding payments are determined using a bingo cage.

Once one of the activities is randomly chosen for payment, subjects are shown a breakdown of their earnings for the session. In addition to their earnings in the chosen activity, they each receive $GTQ20$ as a show-up fee, and $GTQ25$ for completing all seven activities.[10] While payments are being prepared, subjects fill out a short demographic questionnaire.

Subjects were recruited via ORSEE (Greiner, 2015), and were primarily undergraduates at Universidad Francisco Marroquin. Eleven experimental sessions lasting approximately 1.5 hours were conducted, for a total of 132 subjects (60 female). Sessions took place at *Centro Vernon Smith de Economia Experimental*, and each computer station had privacy dividers. Subjects interacted exclusively through the computer interface, which was programmed in z-Tree (Fischbacher, 2007). Participants earned an average of $GTQ116.02 \approx USD14.86$.[11]

[8] In five of the eleven sessions, payment for this belief elicitation was modeled after Niederle and Vesterlund (2007), with each correct answer resulting in a certain payment of $GTQ5$. We prefer our payment scheme, under which payment for belief elicitation only occurs if randomly chosen for payment, as it avoids wealth effects. However, we decided to employ the method of Niederle and Vesterlund (2007) in a subset of our sessions to increase the comparability of our data. In our regression analysis we control for this, are results are robust.

[9] Subjects are asked to make ten choices between two lotteries. Option A always involves two possible payments: $GTQ455.85$ or $GTQ557.15$. Option B always involves a payment of either $GTQ0$ or $GTQ14518.85$. The first choice assigns a 10% chance to the higher payoff in both lotteries. Each subsequent choice increases this by 10%, so that the last choice involves a choice between $GTQ55$ and $GTQ145$. See Appendix A for the instructions, as well as a figure which shows how this risk elicitation task was presented to subjects.

[10] Since payments of $GTQ0$ are possible under a tournament payment scheme, we added the completion payment of $GTQ25 \approx USD3.25$, so that subject earnings would not be too low on average.

[11] As a reference, student workers in the library earned $GTQ24$ per hour and lunch could be purchased on campus for $GTQ25$.

4 Results

We begin our analysis by focusing on our primary research question: how does relative performance feedback affect subsequent tournament performance? Our measure of performance is the number of correct answers submitted during the real-effort task, and we focus on the difference in performance between the two tournaments in which all group members were simultaneously competing. That is, our analysis focuses on the difference in performance after receiving feedback (Activity 6), and the first tournament subjects participated in (Activity 2). On average, feedback has no effect on performance (sign test, $p \approx 1$). However, this result is driven by opposing gender effects in which feedback decreases the performance of men (sign test, $p = 0.0479$) and increases the performance of women (sign test, $p = 0.0479$). This is illustrated in Figure 1, which contains a histogram by gender. The gender difference in the performance response to feedback is significant (Mann-Whitney, $p = 0.016$).

While the overall effect of feedback is of interest, the rank itself has informational content in a winner-take-all environment. A subject who is told they previously held the highest rank is less likely to need to increase their performance in order to win a subsequent contest, and may become complacent. On the other hand, when a subject is told they did not win the previous tournament, they know they must likely increase performance, perhaps dramatically, in order to win.

In addition, rank may provide information that affects beliefs and self image. Roberts and Nolen-Hoeksema (1989) finds that feedback has a larger effect on the self-assessments of women, and that women reacted differently to positive and negative feedback, while men did not. In addition, Niederle and Vesterlund (2007) suggest that "...women, more than men, may view a negative signal as indicative of their self-worth rather than simply their one-time performance on a task." With this in mind, we assess whether or not the effect of feedback differs by whether or not a subject won the previous tournament. Figure 2 summarizes the performance response to feedback by rank and gender.

For men, the change in tournament performance does not depend on whether or not their earlier rank was equal to one (Mann-Whitney, $p = 0.872$). This does not hold for women: the change in tournament performance is higher when their previous rank is less than one (Mann-Whitney, $p = 0.0026$). In fact, women with $Rank = 1$ do not increase their performance (sign test, $p = 0.629$), while women with $Rank \in \{2, 3, 4\}$ do (sign test, $p = 0.0029$). Table 1 contains summary statistics of performance response, as well as the results of Mann-Whitney tests of gender differences in performance response broken down by whether the subject was the earlier victor. Note that there is no significant gender difference when, $Rank = 1$ ($p = 0.770$), and a substantial gender difference when $Rank \in \{2, 3, 4\}$ ($p = 0.001$).[12]

To give a clear view of how the effect of feedback differs by rank, Figure 3 presents the relevant distributions broken down by rank and gender. These distributions make it clear that the differences in the response to feedback are driven by $Rank = 2$ and $Rank = 4$. To further investigate this, we report regression results in which we control for ranks, with a $Rank = 1$ as the excluded category. We also control for performance under the piece rate payment scheme, $Activity_1$ as a measure of ability, the difference in performance under the piece-rate and initial tournament, $Activity_2 - Activity_1$, as a measure of how a subject's performance responds to competition, and whether or not belief elicitation over ranks was incentivized with certainty or not ($BeliefPay$). In

[12]Table 1 in Appendix B contains these summary statistics and non-parametric tests for each possible rank.

a second specification, we include a dummy variable for gender ($Female$). In a third specification we interact gender with each included rank. In a final specification, we control for three additional variables. First is competitiveness as measured in Activity 3 ($Competitiveness_3$).[13] Second is overconfidence, which we define as the difference between the believed and observed rank in the initial tournament ($Overconfidence$). Third is risk preferences, as measured by the number of times a subject choose the relatively "safe" lottery in the risk elicitation task ($SafeCoices$). In all specifications, standard errors are clustered at the group level. Table 2 contains the results.[14] Consistent with our non-parametric tests, no rank is statistically significant when, in specification 1, we do not control for gender effects. The second specification shows that there is a gender difference, but the ranks are still not significant, except for $Rank = 4$ (marginally). Once gender is interacted with the ranks, in specification 3, note that all ranks are statistically significant for women, but not for men (although only marginally so for $Rank = 3$). Controlling for competitiveness, overconfidence and risk aversion does not change the results from the second specification, and not of the coefficients on these variables are significant.

These results are striking. When a women is told that she ranked second in the previous tournament, her performance increases dramatically more than when here previous rank was first. This is consistent with a belief that, with additional effort, she could improve her rank and win the tournament. Similar feedback has no effect on a man's performance. Even more striking, this increase in performance is not diminished when $Rank = 4$ (t-test, $p = 0.649$). When a woman is told that she ranked last in the previous tournament, her probability of winning the subsequent contest is quite low, but does not reduce her performance relative to the case of $Rank = 2$. This is suggestive evidence that beliefs over the probability of winning are the not primary drivers of women's performance improvement, and is consistent with the idea that woman may internalize negative feedback. The novelty of the result is that women are likely to respond to such feedback by dramatically increasing their effort, even when they are not likely to financially gain from doing so. That is, rather than being discouraged by the negative feedback, women increase their effort and performance.

4.1 Competitiveness, overconfidence and risk attitudes

Specification 3 from Table 2 demonstrates that competitiveness, overconfidence and risk aversion do not explain our main results (F-test, $p = 0.759$). However, as each of these are the subjects of significant literatures in their own right, we now discuss each of them in turn, and relate our results to the literature.

We now turn attention to the two measures of competitiveness. The first measure, from Activity 3, is the weight a subject places on the tournament payment scheme knowing that payments will be determined by comparing her subsequent performance in the real-effort task to the previous tournament performance of her group members. Our second measure is also the weight a subject places on the tournament payment scheme, but in this case, the chosen payment scheme is to be applied to the past performance of the group in Activity 1. Summary statistics by gender, as well

[13] Our results are robust to using the measure of competitiveness from Activity 4. Since the measure from Activity 3 involved engaging in a tournament, we opted to use it here.

[14] Table A2 in Appendix B contains regression results where ranks 2, 3 and 4 are collapsed into a single dummy variable.

as the results of non-parametric tests of gender differences are found in Table 3.[15]

Under both measures, men are more competitive on average. However, the differences are not significant at conventional levels. It is important to note that our measures of competitiveness differs from that of Niederle and Vesterlund (2007), in that we allow subjects to choose a linear combination of payment schemes, rather than restricting to a binary choice. It is also important to note that, while most of the literature finds a significant gender difference, we are not the first who fail to do so. Boschini et al. (2019) find no difference for a verbal task, and only weak evidence for a math task. Cárdenas et al. (2012) finds no gender differences among children aged 9–12 in Colombia.[16] Price (2010), using the same software and instructions as Niederle and Vesterlund (2007), also found no gender difference. Price finds that the lack of replication is primarily due to a difference in (over)confidence between subject pools.[17]

With this in mind, we now turn attention to overconfidence, defined as the difference between a subjects believed rank and observed rank, so that a small number corresponds to a higher level of overconfidence. We observe two measures of overconfidence. The first corresponds to performance under the piece-rate payment scheme, while the second corresponds to performance in the initial tournament. Table 3 contains summary statistics of both of these measures, elicited beliefs over ranks, as well as the results of non-parametric tests of underlying gender difference.[18] Note that men are slightly more overconfident, on average, under both payment schemes, although the magnitude of the difference is small in both cases. Further, this difference is insignificant for piece-rate performance, and only marginally significant for tournament performance. Thus, as in Price (2010), the lack of a significant gender difference in competitiveness could be explained by relatively small differences in overconfidence.

We now turn attention to risk attitudes, as measured in Activity 7. For simplicity, we use the number of times a subject chose Option A in Activity 7 as our measure of risk aversion, since Option A was the relatively safe lottery. By relying on the number of safe choices, we avoid problems related to inconsistent answers in this task.[19] Table 3 contains summary statistics of this measure, as well as a non-parametric test of gender differences.[20] Consistent with the existing literature (see Croson and Gneezy (2009); Eckel and Grossman (2008) for an excellent review of the literature), we find that men are less risk averse than woman.

5 Conclusion

This paper studies the effect of relative performance feedback in competitive environments. Our main focus is on the effect of the *content* of such feedback on subsequent performance. Since it has been suggested that women internalize feedback, especially negative feedback, more than men, we

[15] Kernel density plots of this measure by gender can be found in Figures A2 and A3 in Appendix C.

[16] Results with same aged children in Sweden are mixed. Whereas boys are more likely to choose to compete in general, girls tend to improve their performance in response to competition more than bots in some tasks.

[17] Gillen et al. (2019) argues that overconfidence is a primary driving factor of observed differences in competitiveness, when correcting for measurement issues.

[18] Figures A4 and A5 in Appendix C show the distributions of these measures by gender.

[19] If a subject switches back and forth between Option A and Option B multiple times, the data is difficult to interpret. Subjects could be confused or could be indifferent between the lotteries across multiple rows.

[20] Figure A6 in Appendix C contains the distributions of this measure of risk attitudes by gender.

also focus on gender differences.[21]

We find that relative performance feedback has no overall impact on performance. However, this is the result of opposing gender effects. Men, on average reduce slightly reduce their performance after receiving feedback, while women increase their performance. Importantly, only women have differential responses to the content of the received feedback. Specifically, women who learn that they were not the highest ranked person in an initial competition are the only ones who respond by increasing their subsequent performance.

Since the use of relative performance feedback is ubiquitous in the workplace, these findings have important implications. The finding that many women increased their performance after relative feedback demonstrates that there may be benefits to implementing similar practices in the workplace. It may also mitigate concerns that women internalize negative feedback; in fact, men in this experiment may have internalized negative feedback more so than women. It could be that relative performance feedback is an effective mechanism for reducing gender gaps in the workplace. However, it is important to compare the type of feedback that was given in this experiment to that given in the workplace. The feedback in the experiment was clear, objective, and from a neutral source. Feedback in the workplace may be substantially different; it may come from a superior in a bad mood or it may be equivocal and biased. Future research could examine how the gender and status of the feedback-giver influences subsequent performance.

One concern with the use of relative feedback is that the feedback could have detrimental effects on the mental health of employees. One promising avenue for future research is to examine the effect of relative performance feedback on job satisfaction, self-assessments, and subsequent performance in an environment in which the responses to repeated feedback are observed.

[21] See e.g., Roberts and Nolen-Hoeksema (1989).

References

Azmat, G., Bagues, M., Cabrales, A., and Iriberri, N. (2019). What you don't know... can't hurt you? a natural field experiment on relative performance feedback in higher education. *Management Science*, 65(8):3449–3947.

Azmat, G. and Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7-8):435–452.

Azmat, G. and Iriberri, N. (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics & Management Strategy*, 25(1):77–110.

Baldwin, M., Granzberg, A., Pippus, L., and Pritchard, E. (2003). Cued activation of relational schemas: Self-evaluation and gender effects. *Canadian Journal of Behavioural Science*, 35(2):153.

Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment.

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment.

Boschini, A., Dreber, A., von Essen, E., Muren, A., and Ranehill, E. (2019). Gender, risk preferences and willingness to compete in a random sample of the swedish population. *Journal of Behavioral and Experimental Economics*, 83:101467.

Cárdenas, J.-C., Dreber, A., Von Essen, E., and Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in colombia and sweden. *Journal of Economic Behavior & Organization*, 83(1):11–23.

Charness, G., Masclet, D., and Villeval, M. C. (2014). The dark side of competition for status. *Management Science*, 60(1):38–55.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2):448–74.

Eckel, C. and Grossman, P. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of experimental economics results*, 1:1061–1073.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–38.

Eriksson, T., Poulsen, A., and Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688.

Ertac, S. (2011). Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, 80(3):532–545.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Garcia, S. M., Tor, A., and Schiff, T. M. (2013). The psychology of competition: A social comparison perspective. *Perspectives on Psychological Science*, 8(6):634–650.

Gill, D., Kissová, Z., Lee, J., and Prowse, V. (2019). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, 65(2):494–507.

Gillen, B., Snowberg, E., and Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4):1826–1863.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Hannan, L., Krishnan, R., and Newman, A. (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *Accounting Review*, 83(4):893–913.

Holt, C. and Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Huang, N., Burtch, G., Gu, B., Hong, Y., Liang, C., Wang, K., Fu, D., and Yang, B. (2018). Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science*, 65(1):327–345.

Krähmer, D. (2007). Equilibrium learning in simple contests. *Games and Economic Behavior*, 59(1):105–131.

Kuhnen, C. and Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1):94–113.

Niederle, M. (2016). Gender. In Kagel, J. and Roth, A., editors, *Handbook of Experimental Economics*, page 481–553. Princeton University Press, 2 edition.

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

Prendergast, C. (1999). The provision of incentives in firms. *Journal of economic literature*, 37(1):7–63.

Price, C. (2010). Do women shy away from competition? do men compete too much?: A (failed) replication. *Economics Bulletin*, 40(2):1538–1547.

Roberts, T.-A. and Nolen-Hoeksema, S. (1989). Sex differences in reactions to evaluative feedback. *Sex Roles*, 21(11-12):725–747.

Saccardo, S., Pietrasz, A., and Gneezy, U. (2017). On the size of the gender difference in competitiveness. *Management Science*, 64(4):1541–1554.

Tesser, A. and Campbell, J. (1980). Self-definition: The impact of the relative performance and similarity of others. *Social Psychology Quarterly*, 43:341–347.

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

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, 110(2):337–61.

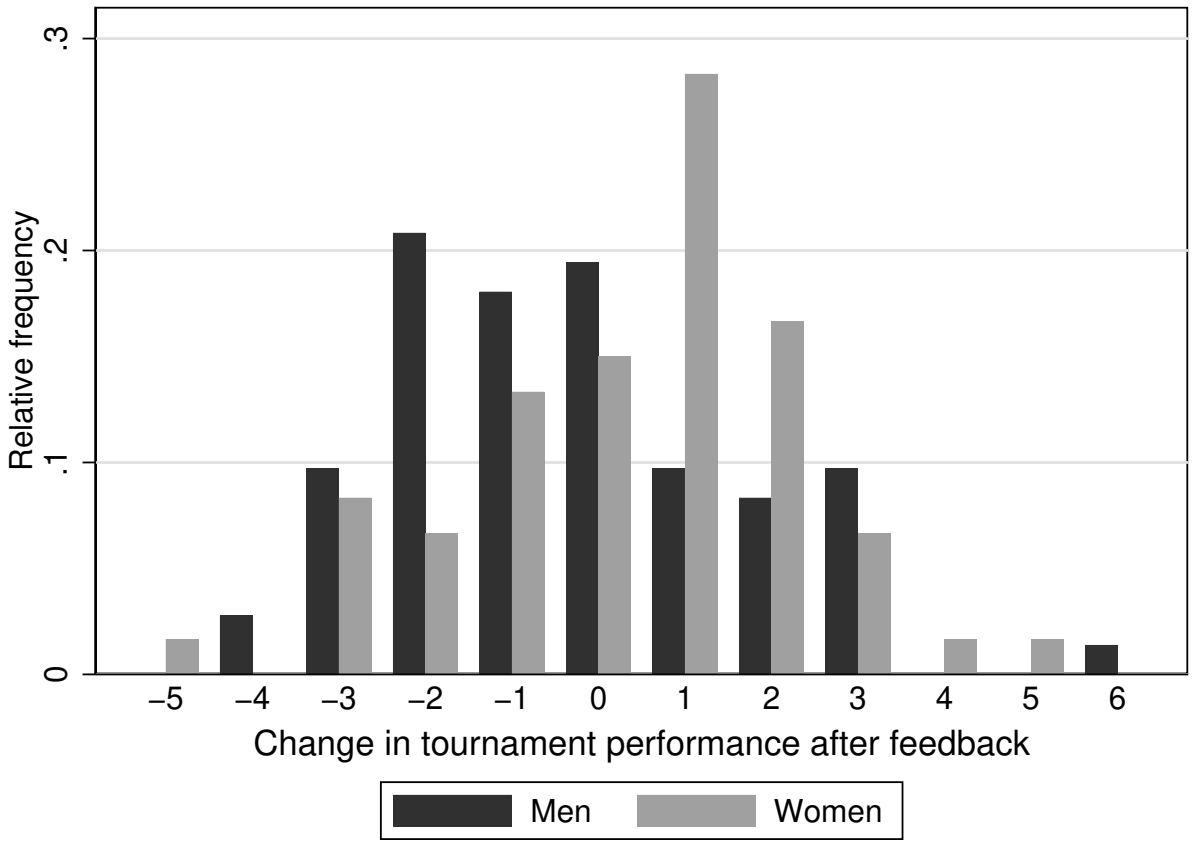


Figure 1: Change in tournament performance after feedback

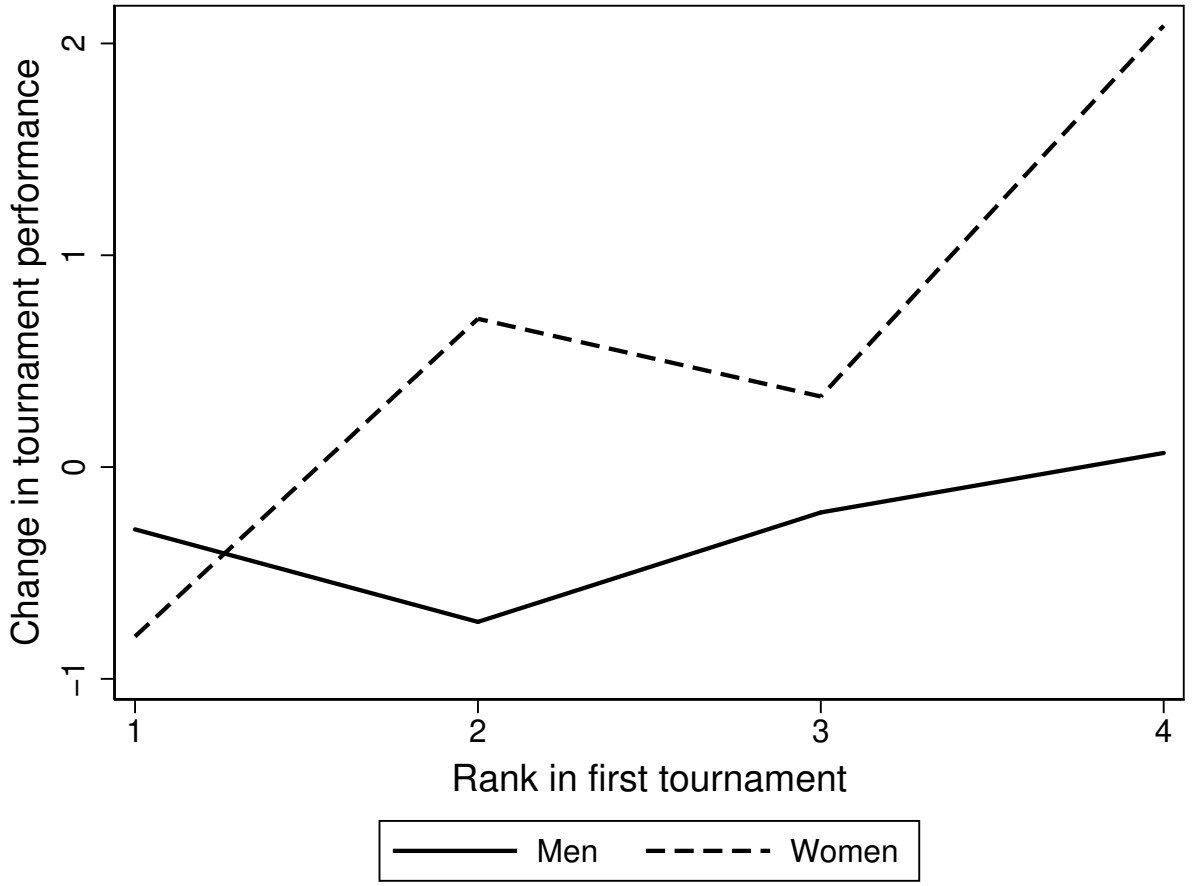


Figure 2: Change in tournament performance after feedback by rank and gender

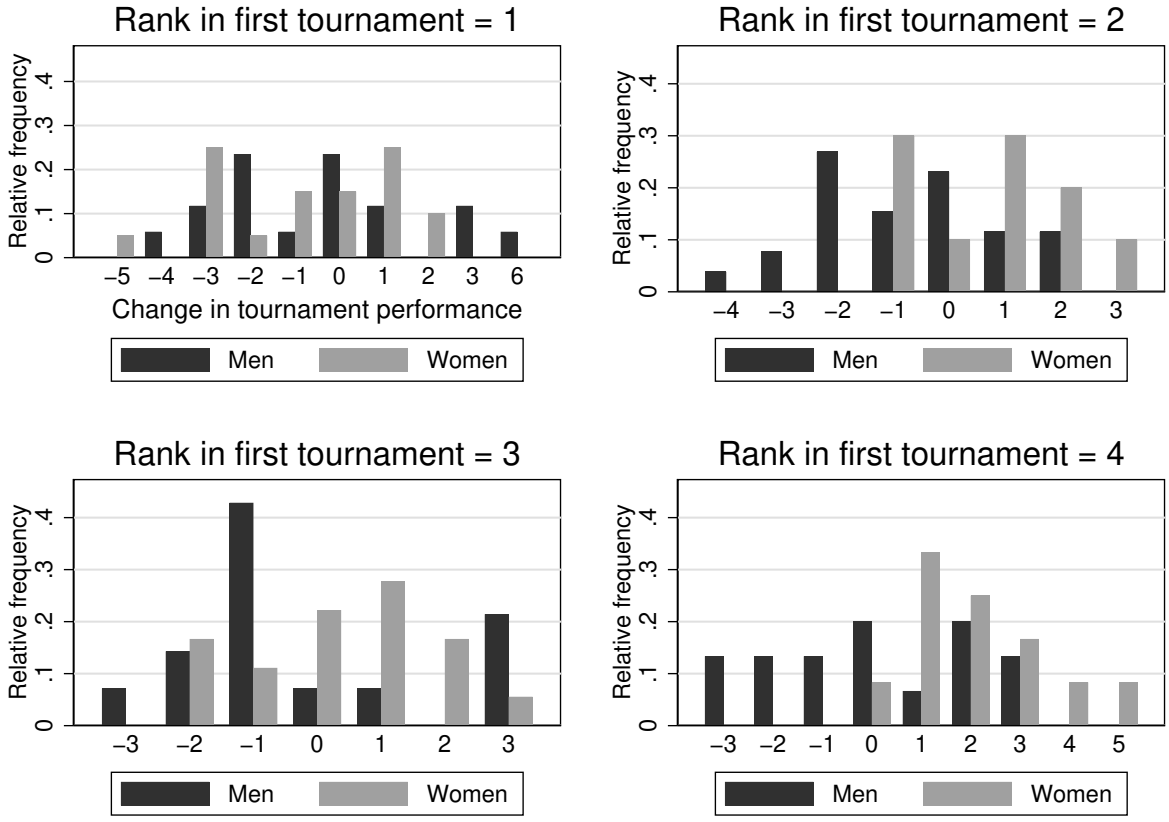


Figure 3: Change in tournament performance after feedback by rank and gender

Table 1: Change in tournament performance after feedback

	Men	Women	p-value
All ranks	-0.361	0.367	0.016
	(2.016)	(1.931)	
$Rank = 1$	-0.294	-0.800	0.770
	(2.568)	(2.016)	
$Rank \in \{2, 3, 4\}$	-0.382	0.950	0.001
	(1.841)	(1.616)	
$Rank = 2$	-0.731	0.700	0.025
	(1.638)	(1.418)	
$Rank = 3$	-0.214	0.333	0.280
	(1.968)	(1.495)	
$Rank = 4$	0.067	2.083	0.018
	(2.052)	(1.443)	

Notes: Table contains means with standard deviations in parentheses. p-values are for two-tailed Mann-Whitney U-tests.

Table 2: Regression results for change in performance after feedback

	(1)	(2)	(3)	(4)
$Rank = 2$	0.306	0.505	-0.443	-0.367
	(0.523)	(0.500)	(0.757)	(0.843)
$Rank = 3$	0.555	0.578	-0.443	-0.372
	(0.557)	(0.515)	(0.748)	(0.959)
$Rank = 4$	1.131	1.278*	-0.012	0.195
	(0.796)	(0.744)	(0.792)	(1.4)
$Ability : Activity_1$	0.040	0.048	0.045	0.063
	(0.131)	(0.134)	(0.127)	(0.147)
$Competition : Activity_2 - Activity_1$	-0.314**	-0.291**	-0.309***	-0.287**
	(0.121)	(0.122)	(0.107)	(0.132)
$BeliefPay$	0.230	0.310	0.358	0.408
	(0.321)	(0.310)	(0.3)	(0.299)
$Female$		0.713**	-0.770	-0.708
		(0.329)	(0.76)	(0.754)
$Rank = 2 \cdot Female$			1.968**	2.044**
			(0.943)	(0.983)
$Rank = 3 \cdot Female$			1.935*	1.963*
			(1.049)	(1.065)
$Rank = 4 \cdot Female$			2.468***	2.437***
			(0.735)	(0.782)
$Competitiveness_3$				-0.002
				(0.008)
$SafeChoices$				-0.118
				(0.109)
$Overconfidence$				0.025
				(0.264)
$Constant$	-0.616	-1.166	-0.322	0.037
	(1.377)	(1.372)	(1.326)	(1.527)
Observations	132	132	132	132
Clusters	33	33	33	33
R^2	0.172	0.202	0.258	0.263
BIC	566.2647	566.3999	571.3374	585.1151
AIC	546.0851	543.3375	539.6265	544.7559

Notes: Dependent variable is the difference in number of correct answers between the two winner take all tournaments, after (Activity 6) - before (Activity 2) receiving feedback.

Standard errors are clustered at the group level. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 3: Summary statistics for competitiveness, overconfidence and risk preferences

	Men	Women	p-value
% to tournament in Activity 3	0.501	0.458	0.315
	(0.262)	(0.288)	
% to tournament in Activity 4	0.369	0.360	0.954
	(0.321)	(0.281)	
Believed piece rate rank	2.208	2.250	0.6110
	(1.006)	(0.876)	
Believed tournament rank	1.806	2.117	0.124
	(0.816)	(1.059)	
Believed - observed piece rate rank	-0.250	-0.033	0.307
	(1.33)	(1.327)	
Believed - observed tournament rank	-0.569	-0.250	0.054
	(1.059)	(1.002)	
Number of safe choices in risk elicitation task	4.875	5.533	0.003
	(1.453)	(1.186)	

Notes: Table contains means with standard deviations in parentheses. p-values are for two-tailed Mann-Whitney U-tests.

A Instructions

This appendix contains the instructions for each of the seven activities, translated from the original Spanish.

A.1 Introduction

SLIDE NUMBER 1

Introduction

- The following instructions will explain how you can earn money. The amount of money that each participant earns may vary considerably depending on the decisions the participant makes.
- Participants will interact only through computers. If anyone talks, makes any kind of signs, looks into other participant's computer, etc., we will terminate the experiment and will ask you to leave without any earnings.

SLIDE NUMBER 2

General overview

- In this experiment you will have to complete seven different activities. Only one of these seven activities – randomly selected – will be taken into account for your payment. Each of these activities has the same probability in being chosen.
- At the end of the experiment, we will put seven balls in a bingo cage and we will remove one of them to determine which of the seven activities will be taken into account for the payment of your earnings.

SLIDE NUMBER 3

- In addition to the payment of the randomly selected activity and your participation payment ($Q20$), you will receive $Q25$ for completing the seven activities.
- The seven activities differ. Several of them ask you to calculate the sum of five two-digit random number for 5 minutes. Before you start any activity, we will give you the opportunity to become familiar with the interface and to practice for five minutes.
- Then, before beginning each activity, we will explain to you how your earnings will be determined in case a particular activity is chosen.

A.2 Activity 1

SLIDE NUMBER 1

Activity 1 - Unitary

- For Activity 1 you will be asked to calculate the sum of five two-digit random numbers. You will have 5 minutes to calculate the correct sum of a series of these problems. You may not use calculator, but you can use a blank sheet of paper and pencil, which will be provided to you.
- You can send your response by clicking on the submit button. Upon entering your response, you will instantly know if is correct or not.

SLIDE NUMBER 2

- If Activity 1 is randomly selected for payment, you will earn $Q5$ for each correct answer you get in 5 minutes. Your earning will not be reduced if your answer is incorrect.
- Upon completing this activity, you will only be able to see the number of correct answer you submitted.
- If you have any questions before beginning, please raise your hand.

A.3 Activity 2

SLIDE NUMBER 1

Activity 2 - Tournament

- As in Activity 1, you will have 5 minutes to calculate the correct sum of a series of five two-digit numbers.
- However, for this Activity, we will form groups (randomly) of four persons and only the person from each group that gets the most correct answers will receive earnings.

SLIDE NUMBER 2

- If this Activity is randomly chosen for payment, the person with the most correct answers will receive $Q20$ for each one. (In case of ties, the winner is whoever entered the last correct answer first). The rest of the people in that group will not receive any kind of payment.
- Upon finishing this activity, you will only see the number of correct answers you submitted.
- If you have any questions before beginning, please raise your hand.

A.4 Activity 3

SLIDE NUMBER 1 **Activity 3 – Combination**

- As in the last two Activities, you will have 5 minutes to calculate the correct sum of a series of five two-digit numbers.
- However, now you will be able to choose the payment scheme for the correct answers you get in this Activity. Specifically, you will be able to choose which percentage of the total payment you want to be by Unitary and which percentage by Tournament:

$$\begin{array}{l} \%\text{ Payment by Unitary } (Q5 \times \text{correct answer}) \\ +\ \underline{\%\text{ Payment by Tournament } (Q20 \times \text{correct answer, if you win})} \\ 100\%\text{ (Total Payment)} \end{array}$$

SLIDE NUMBER 2

- The Payment by Unitary provides $Q5$ for each correct answer.
- The Payment by Tournament provides $Q20$ for each correct answer in this Activity (3), only if you get a higher number of correct answers than the number of correct answers that the other three participants in your group got in Activity 2 (Tournament) – that is, the Activity which you just completed. (In case of ties, the winner is whoever entered the last correct answer first).
- If you do NOT get more correct answers than the other participants in your group in Activity 2 (Tournament), the Payment by Tournament does not yield any payment.

SLIDE NUMBER 3

- If this Activity is randomly chosen for payment, and you get a higher number of correct answers than the number of correct answers that the other three participants in your group got in Activity 2 (Tournament), you will get:

$$\begin{array}{c} Q5 \times (\%\text{ Payment by Unitary}) \ \times (\#\text{ correct answers}) \\ +\ Q20 \times (\%\text{ Payment by Tournament}) \ \times (\#\text{ correct answers}) \end{array}$$

- If this Activity is randomly chosen for payment, and you do NOT get a higher number of correct answers than the number of correct answers that the other three participants in your group got in Activity 2 (Tournament), you will get:

$$Q5 \times (\%\text{ Payment by Unitary}) \times (\#\text{ correct answers})$$

SLIDE NUMBER 4

- On the next screen you will be asked to choose the percentage of the total payment you wish to be by Unitary and the percentage by Tournament. Then you will have 5 minutes to calculate the sums.
- Upon finishing this activity, you will only see the number of correct answers you obtained.
- If you have any questions before beginning, please raise your hand.

A.5 Activity 4

SLIDE NUMBER 1

Activity 4 – Combination II

- In this Activity you will have to make a similar selection as in Activity 3, with the exception that you will not be asked to sum more numbers. This time you will be able to choose the payment scheme for the correct answers you got in Activity 1 (Unitary). Specifically, you will be able to choose which percentage of the total payment you want to be by Unitary and which percentage by Tournament:

$$\begin{array}{l} \%\text{ Payment by Unitary } (Q5 \times \text{correct answer}) \\ + \underline{\%\text{ Payment by Tournament } (Q20 \times \text{correct answer, if you win})} \\ 100\%\text{ (Total Payment)} \end{array}$$

SLIDE NUMBER 2

- The Payment by Unitary provides $Q5$ for each correct answer.
- The Payment by Tournament provides $Q20$ for each correct answer in Activity 1 (Unitary), only if you got a higher number of correct answers than the other participants in your group in Activity 1 (Unitary). (In case of ties, the winner is whoever entered the last correct answer first).
- If you did NOT get more correct answers than the other participants in your group in Activity 1 (Unitary), the Payment by Tournament does not yield any payment.

SLIDE NUMBER 3

- If this Activity is randomly chosen for payment, and you got the highest number of correct answers in your group in Activity 1 (Unitary) in comparison with the other participants in your group, you will get:

$$Q5 \times (\%\text{ Payment by Unitary}) \times (\#\text{ correct answers})$$
$$+\, Q20 \times (\%\text{ Payment by Tournament}) \times (\#\text{ correct answers})$$

- If this Activity is randomly chosen for payment, and you did NOT got the highest number of correct answers in your group in Activity 1 (Unitary) in comparison with the other participants in your group, you will get:

$$Q5 \times (\%\text{ Payment by Unitary}) \times (\#\text{ correct answers})$$

SLIDE NUMBER 4

- On the next screen you will be able to see the number of correct answers that you got in Activity 1 (Unitary). Also, you will be asked to choose the percentage of the total payment you wish to be by Unitary and the percentage by Tournament for your answers, and you will not have to do any more sums in this Activity.
- Upon finishing this activity, you will only see the number of correct answers you obtained.
- If you have any questions before beginning, please raise your hand.

A.6 Activity 5

SLIDE NUMBER 1

Activity 5 – Ranking

- In this Activity you will not have to add up digits. This time we will ask you several questions about what think your position was in the ranking of the members of your group in relation to the number of correct answers.
- If this Activity is randomly chosen for payment, we will choose one of the questions randomly. If for the chosen question you guess your position in the ranking, you get $Q60$. If you do not guess your position, you get $Q10$.
- If you have any questions before beginning, please raise your hand.

A.7 Activity 6

SLIDE NUMBER 1

Activity 6 – Tournament II

- As in Activity 2, you will have 5 minutes to calculate the correct sum of a series of five two-digit numbers.
- For this Activity, you will be participating again with the same people from your group, and only the person of each group that gets the most correct answers will receive earnings.
- Before beginning the activity, you will see your ranking among the members of your group, in relation to the number of correct answers in Activity 2 (Tournament).

SLIDE NUMBER 2

- If this Activity is randomly chosen for payment, the person with the most correct answers will receive $Q20$ for each one. (In case of ties, the winner is whoever entered the last correct answer first. The rest of the persons in that group will not receive any payment.
- Upon finishing this activity, you will only see the number of correct answers you submitted.
- If you have any questions before beginning, please raise your hand.

A.8 Activity 7

SLIDE NUMBER 1

Activity 7 – Options

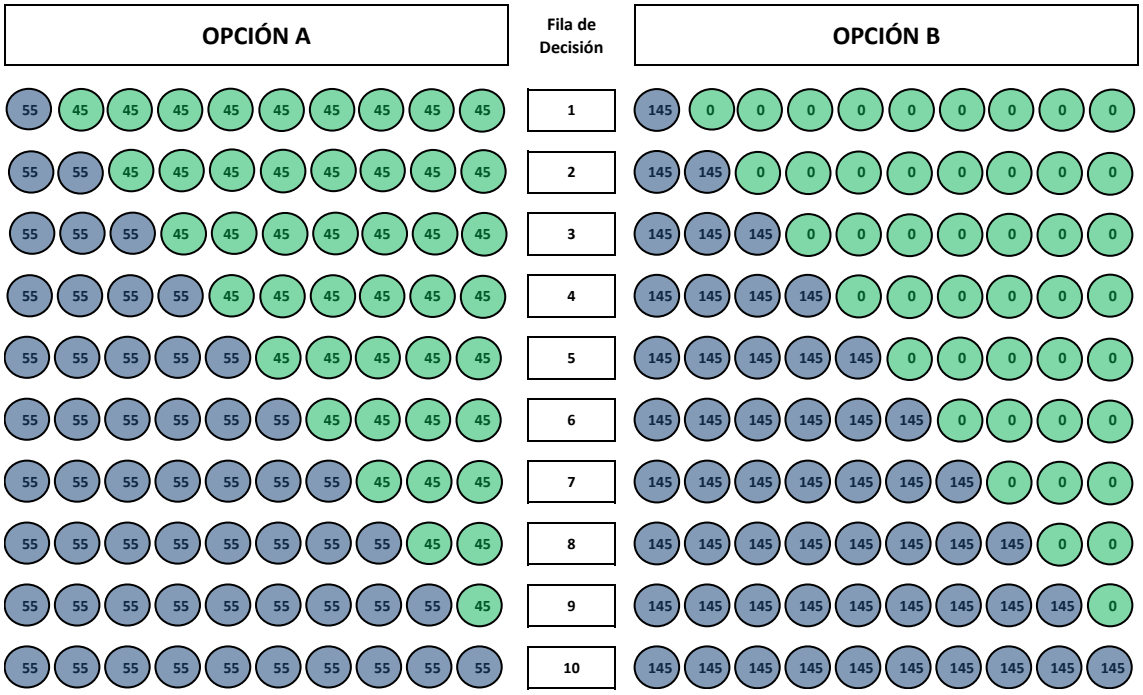
- In this Activity you will not have to add more numbers. This time you will have two options – OPTION A and OPTION B – that will pay money according to the color of the ball – green or blue – that leaves the bingo cage.

If the ball is color...	OPTION A	OPTION B
...BLUE	Q55	Q145
...GREEN	Q45	Q0

- You will see ten decision rows. The number of blue (and green) balls varies according to the decision row. For each row you must choose between OPTION A or OPTION B.

SLIDE NUMBER 2

For this slide, the text was a voice-over of this image:



- In decision row 1 there is 1 blue ball (and 9 green ones); in decision row 2, there are 2 blue balls and 8 green one...and so on until decision row 10, in which there are 10 blue balls and no green. (Note that there will always be 10 balls; what changes is the mix of blue and green balls according to the decision row).
- If this Activity is randomly selected for payment, we will throw a 10-sided die to choose a decision row and then we will draw a ball from the bingo cage to determine your payment, depending on the OPTION that you chose for that row.
- Please note Decision Row 1 (1 blue ball and 9 green balls):
- OPTION A pays $Q55$ if the blue ball comes out; or pays $Q45$ if one (out of 9) of the green balls comes out.
- OPTION B pays $Q145$ if the blue ball comes out; or pays $Q0$ if one (out of 9) of the green balls comes out.
- Note that the probability of drawing a blue ball in this row is 10% (1 blue ball out of 10); the probability of drawing a green ball in this row is 90% (9 green balls out of 10).
- As you move down the number of blue balls increase (and the number of green balls decreases). That is, the probability that the payment is defined by the blue ball (higher payments) increases for both options.

- In fact, for decision row 10 (the last row), there are only blue balls for both option. Therefore, the payment will be, with certainty, the one of the blue ball (100% probability). That is, for this row, you must choose between $Q55$ (OPTION A) and $Q145$ (OPTION B).

SLIDE NUMBER 3

- If Activity 7 is randomly chosen for payment, your earnings will be determined as follows:
- First we will throw a ten-sided die to choose the decision row that will determine your earnings. (Naturally, each decision row has the same probability of being chosen).
- Then, we will put the number of blue and green balls corresponding to the particular decision row in the bingo cage. We will draw a ball from the bingo cage and your earnings will be determined according to the color of the chosen ball and the option that you selected in the decision row.

SLIDE NUMBER 4

For this slide, the text was again a voice-over of the image of the decision rows (see slide 2 from this activity).

- Here is an example. Suppose that your earnings are determined according to this Activity, and that the number of the die is 7. This means that row 7 will determine your earnings and we will put 7 blue balls and 3 green balls in the bingo cage.
- If the bingo cage chooses a blue ball and for row 7 you chose OPTION A, your earnings would be $Q55$. If for row 7 you chose OPTION B, your earnings would be $Q145$.
- If the bingo cage chooses a green ball and for row 7 you chose OPTION A, your earnings would be $Q45$. If for row 7 you chose OPTION B, your earnings would be $Q0$.

SLIDE NUMBER 5

- On the next screen you will see 10 decision rows and you will have to select the option you prefer for each row.
- If you have any questions before beginning, please raise your hand.

B Additional tables

Table A1: Summary statistics of performance during the real-effort task

	Men	Women	p-value
Activity 0: Practice	6.514	6.667	0.694
	(2.415)	(2.653)	
Activity 1: Piece rate	8.444	8.500	0.991
	(2.69)	(2.703)	
Activity 2: Tournament	9.639	9.367	0.644
	(3.186)	(2.957)	
Activity 3: Chosen mixture	9.292	9.217	0.825
	(3.609)	(2.946)	
Activity 6: Tournament after feedback	9.278	9.733	0.141
	(3.758)	(2.414)	

Notes: Table contains means with standard deviations in parentheses. p-values are for two-tailed Mann-Whitney U-tests.

Table A2: Regression results for change in performance after feedback

	(1)	(2)	(3)
$Rank \in \{2,3,4\}$	0.367	-0.551	-0.608
	(0.523)	(0.709)	(0.792)
$Ability : Activity_1$	-0.023	-0.002	0.011
	(0.105)	(0.101)	(0.109)
$Competition : Activity_2 - Activity_1$	-0.388***	-0.378***	-0.369**
	(0.094)	(0.080)	(0.086)
$BeliefPay$	0.216	0.332	0.354
	(0.314)	(0.290)	(0.282)
$Female$		-0.820	-0.778
		(0.776)	(0.777)
$Rank \in \{2,3,4\} \cdot Female$		2.113**	2.195**
		(0.832)	(0.867)
$Competitiveness_3$			-0.001
			(0.008)
$SafeChoices$			-0.107
			(0.102)
$Overconfidence$			-0.099
			(0.173)
$Constant$	0.190	0.331	0.754
	(1.063)	(1.115)	(1.059)
Observations	132	132	132
Clusters	33	33	33
R^2	0.159	0.2430	0.2496
BIC	558.5878	554.4886	567.9826
AIC	544.1738	534.309	539.1545

Notes: Dependent variable is the difference in number of correct answers between the two winner take all tournaments, after (activity 6) - before (activity 2) receiving feedback.
Standard errors are clustered at the group level. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

C Additional figures

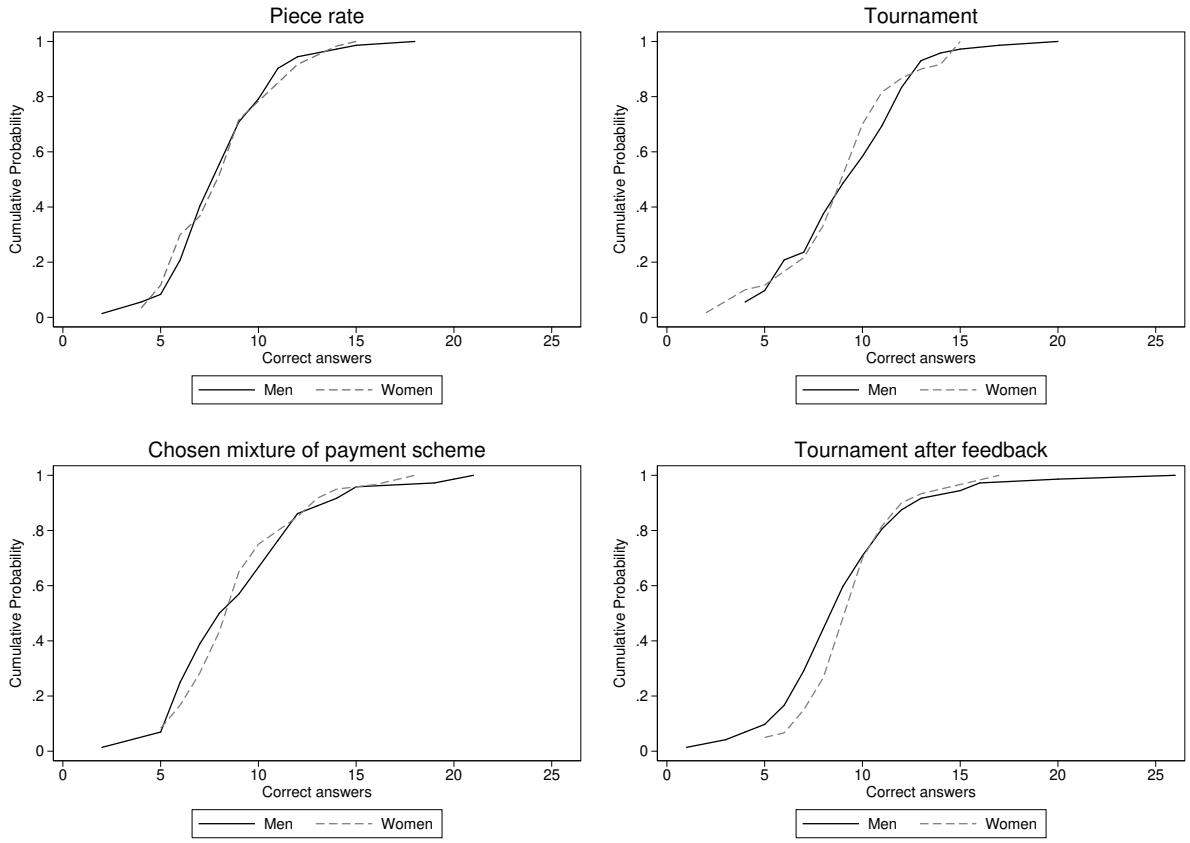


Figure A1: CDFs of performance on the real-effort task by gender

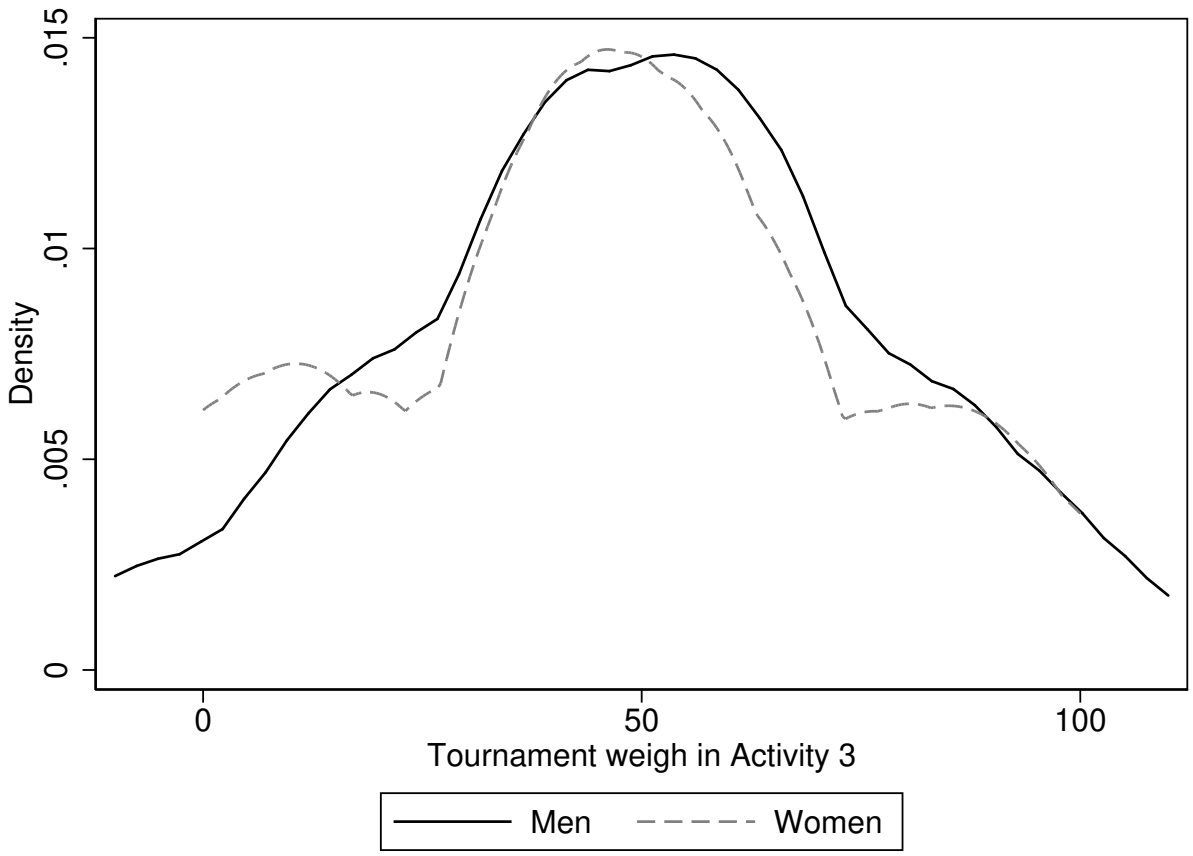


Figure A2: Competitiveness as measured by tournament weight in activity 3

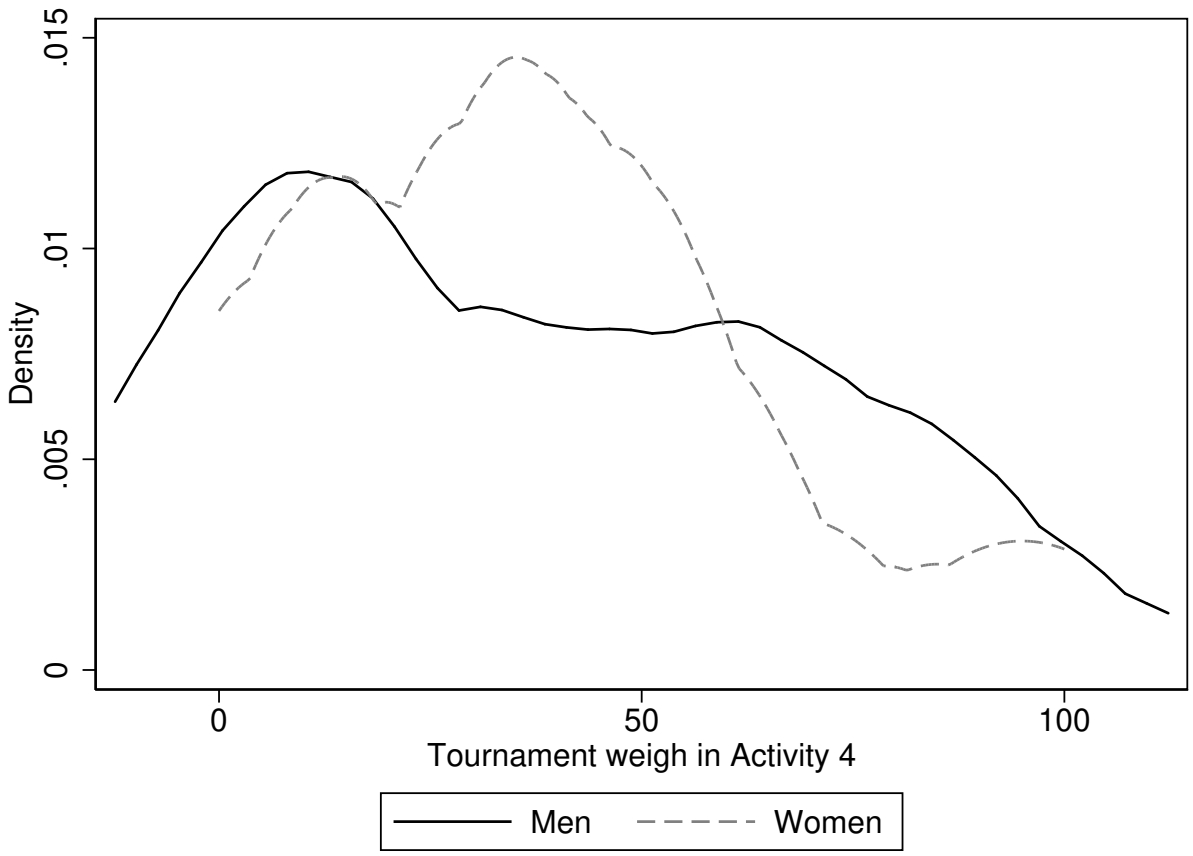


Figure A3: Competitiveness as measured by tournament weight in activity 4

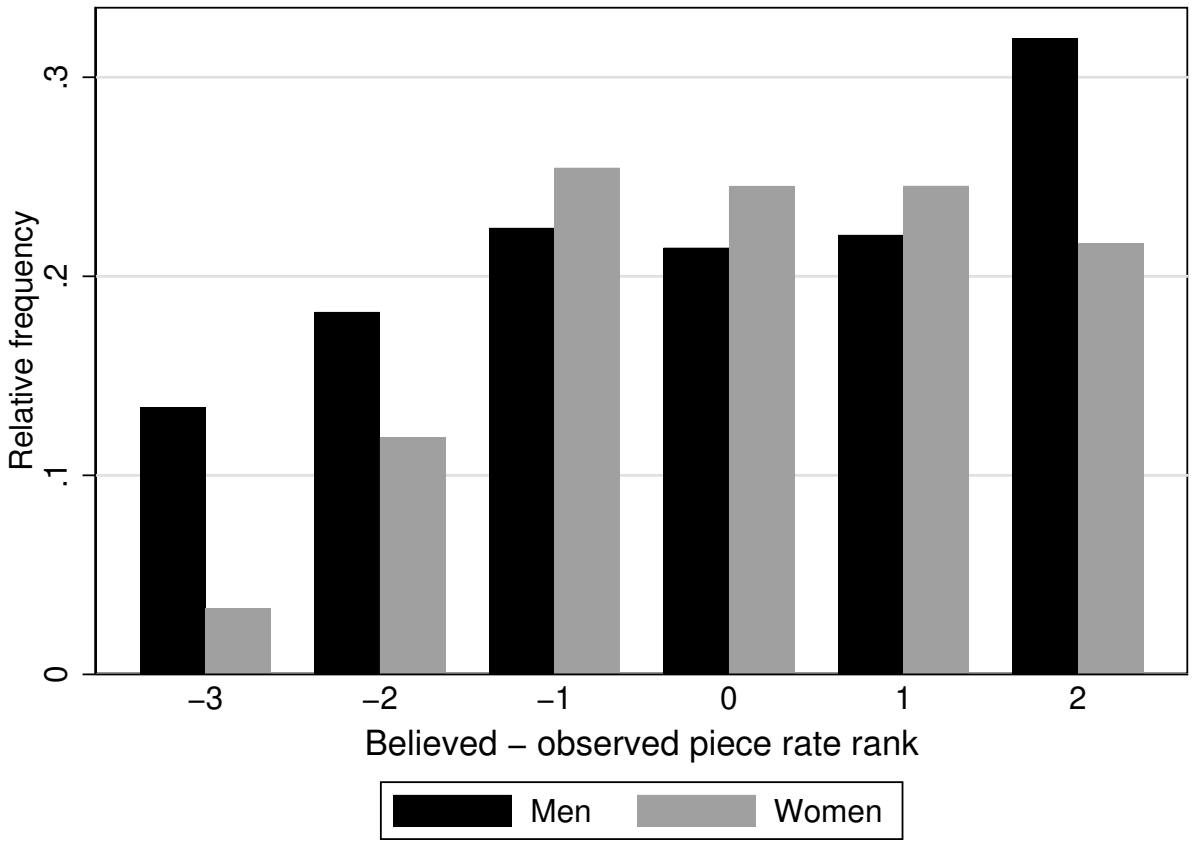


Figure A4: Overconfidence over piece-rate performance

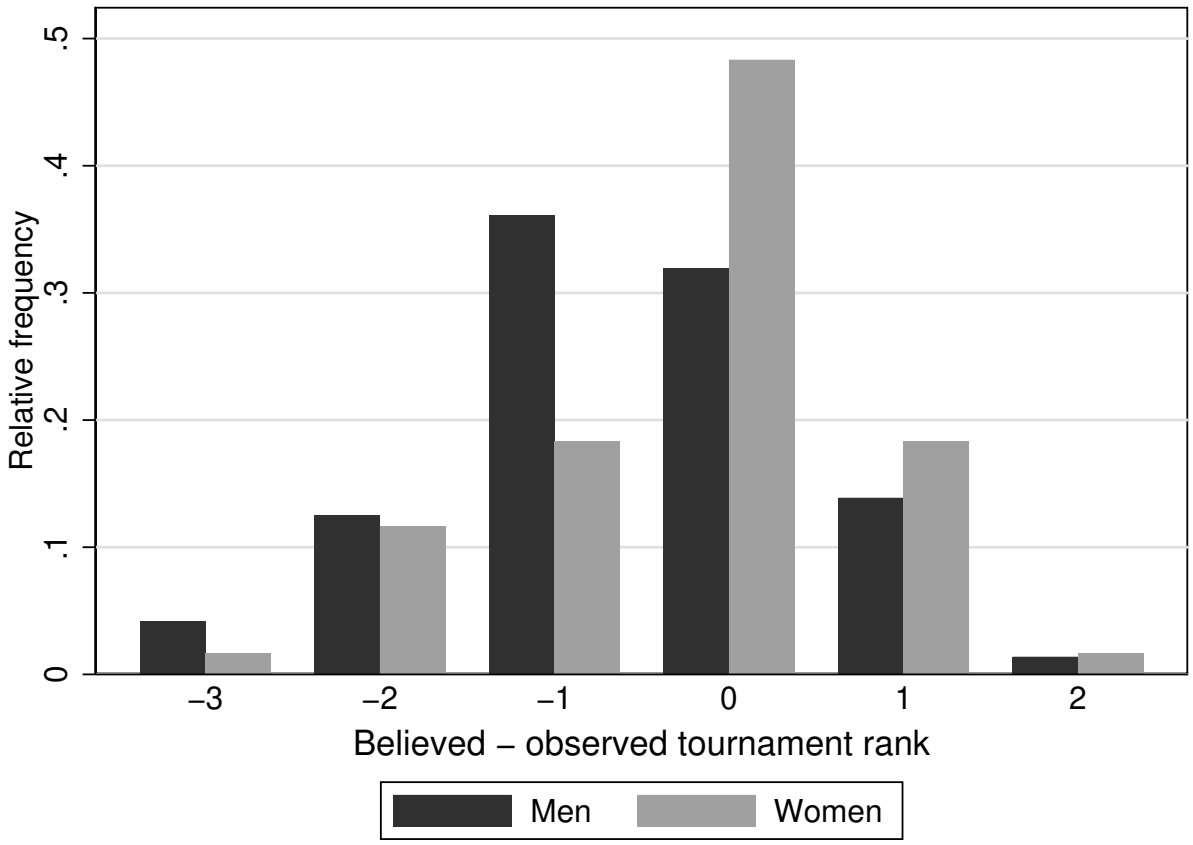


Figure A5: Overconfidence over tournament performance

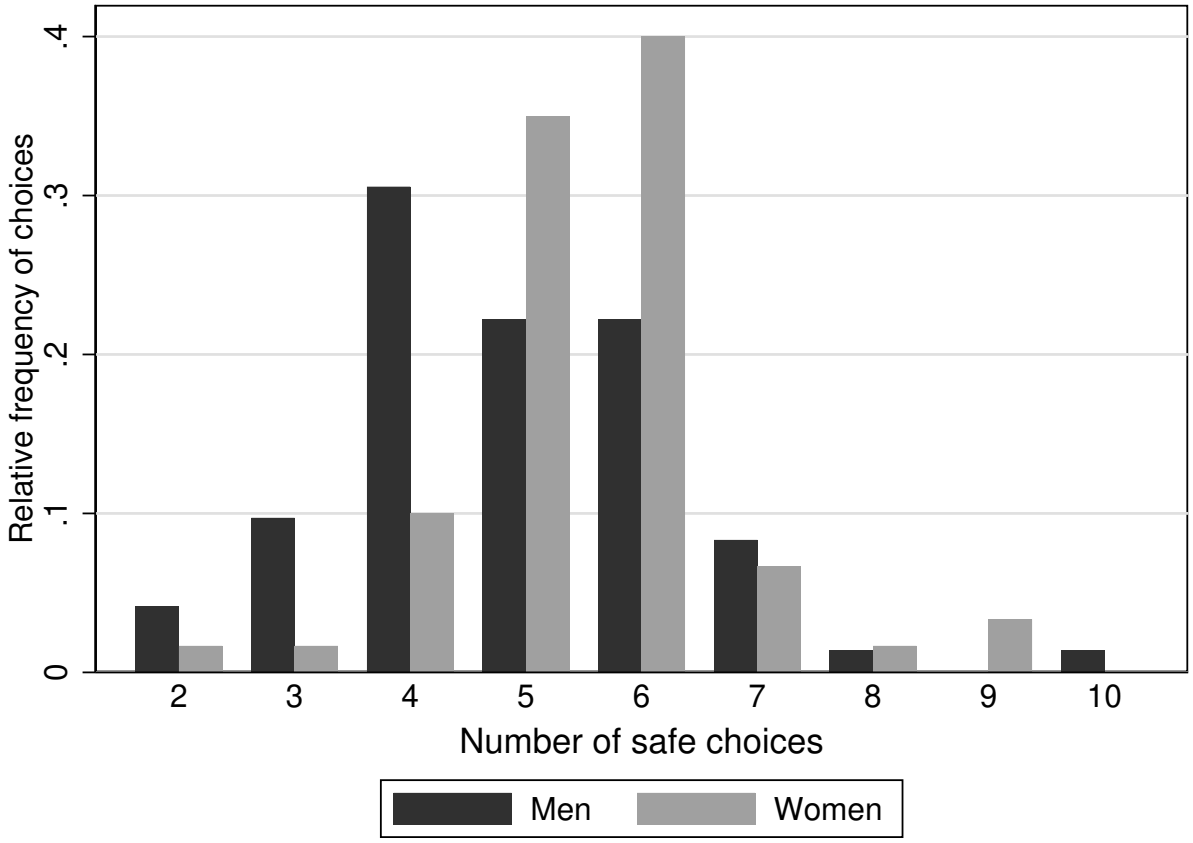


Figure A6: Number of safe choices in the risk elicitation task