The effect of losing a competition: the role of gender, unfairness and feedback

Pre-Analysis Plan

G. Pfeifer S. Piasenti M. Valente R. van Veldhuizen

Index

1 Summary

Using an online experiment we investigate whether there are gender differences in willingness to compete again after winning/losing a competition in both fair and unfair tournaments. This setting closely resembles real-life job competition where participants (i) apply to a fair or possibly unfair competition, (ii) fail, knowing or not whether they had unfair (dis)advantages, and (iii) decide to re-apply or not. By estimating individual-level (covariate-specific) treatment effects with machine learning techniques, we will understand which men and women are driving the gender gap. Thus, our policy implications (such as the introduction of gender quotas) can be, for the first time, tailored to a specific group of the population.

2 Motivation, Related Literature and Contribution

Women currently hold only 5.4% of CEO positions at S&P 500 companies (Catalyst, 2019). Reducing the gender disparity at the top may be desirable not only due to fairness considerations, but may also increase overall productivity (Hoogendoorn *et al.*, 2013). However, in order to eliminate or reduce such gender differences, it is important to understand their causes. Traditional explanations focus on discrimination, differences in gender roles and occupations, and a number of other factors (see e.g., Goldin, 2014; or Blau & Kahn, 2017 for overviews).

A series of studies have linked these differences in the labor market to gender differences in economic preferences and psychological traits, including risk aversion, self-confidence and altruism (see e.g., Croson & Gneezy, 2009 for a review). Perhaps most prominent among these are a number of studies that show that men and women differ in their tendency to seek out competitive environments (Niederle & Vesterlund, 2007). Since prestigious and lucrative jobs are often embedded in competitive environments, this gender difference in willingness to compete may explain some gender differences in labor market outcomes (Buser *et al.*, 2014). Buser & Yuan (2019) showed that men and women may not just differ in their willingness to enter competitive environments, but may also respond differently to adversity: men take adversity in their stride more than women, who instead take a competition loss a signal not to enter future competitions. In addition, Buser *et al.* (2021a) find inconclusive results for gender differences in the willingness to entry competition when individuals know that competition is unfair and competitors are weaker. On the one hand, women seem less likely to exploit unfair advantages. On the other hand, women seem to entry competition more often when they know their competitor will be weaker. We shed new light on attitudes towards competition and unfairness. In particular, we focus on a repeated competition setting and we study the effect of losing a competition when, realistically, individuals do not know whether they won or lost the tournament undeservedly.

Our research project studies attitudes towards (re-)competition through new lenses of perceived unfairness and feedback on the nature of the competition. Many of us have experienced situations where we lost out on a job for reasons we perceived as unfair, for example because of perceived discrimination due to the hiring committee already having its own pre-selected favorite (e.g., an internal candidate). How do we respond to losing out on a job (i.e., losing a competition) in such cases? Will we be less discouraged to try again if we felt that we only lost due to procedural fairness? Or will we be more discouraged due to being convinced that the system is rigged against us? And will the answers to these questions be gender-specific, similar to the response to losing fair competitions?
In summary, our study aims at answering three questions:

1. Do men and women differ in how they respond to losing or winning a competition?

2. Do these gender differences increase or decrease in presence of unfair conditions?

3. What are the mechanisms and characteristics driving any of the observed differences?

To answer these questions, we run an online experiment on the platform Prolific.co. Our study contributes to fill at least three knowledge gaps. First, we examine the understudied question of the impact of perceived unfairness on dropping out of competitions. Our work relates to the literature studying affirmative actions such as quotas or preferential treatments (see, e.g., Balafoutas & Sutter, 2010), with the main difference that we study willingness to compete conditional on losing/winning a competition where unfairness is not gender specific.
Second, we explore gender-specific responses to unfairness. Existing literature shows the presence of a gender gap in unfairness perception (Solnick, 2001; Slonim *et al.*, 1998), and that this difference is likely due to nurture more than nature as it is observed, e.g., for the U.S. but not for Germany and South Korea (Maxwell *et al.*, 2009). Third, we explore treatment effect heterogeneity and relax the assumption that treatment effects are the same among individual with different characteristics. This is important because policy implications of our analysis can be tailored to a specific subgroup of the general population. In order to handle the high-dimensional nature of the collected data, we perform statistical analyses using machine learning techniques.

3 Experimental Design

In our experiment participants are randomly assigned to one of three treatments. Each treatment consists of three effort tasks (plus an intermediate task in which we elicit beliefs about own performances) and concludes with a questionnaire. Treatments are explained in detail below.

3.1 Neutral Treatment

Each player plays three main tasks of 90 seconds each. Only one randomly chosen task will be payoff-relevant. In task 1, participants perform a real effort task in which they have to count the number of zeros (0) in ten tables consisting of zeros (0) and ones (1) (Apicella *et al.*, 2017). They are paid according to a piece rate that pays 0.15 pounds per table they solve correctly. In task 2 participants work on the same task, but are paid according to a tournament rate which pays 0.30 pounds if the participant's score (which is the number of tables they solve correctly) exceeds the score of another randomly selected player who has already played the task. Before task 3 is played, we ask players to consider their performance in task 2 and guess their rank compared to other 100 participants in task 2. We call this subtask the "Guessing task". This subtask is incentivized. We pay a base payment of 0.50 pounds, with a penalty of 0.02 pounds times the absolute difference between the true rank and the stated (guessed) rank.
Subsequently, participants are given a feedback, i.e., "you won/lost in the tournament". After having received the feedback, they are given the possibility to update their estimated rank. In task 3, participants work on the same task but before that, they can choose between the piece rate and tournament payment. If the latter is chosen, participants' scores are matched with the scores of other players who already played the task (different from the opponent in task 2).

3.2 Unfair Treatment

It is analogous to the "Neutral Treatment" except that before completing task 2, participants are told that in 50% of the cases the winner of the tournament will be the one with the higher score (i.e.: the "real" winner) while in the remaining 50% of the cases the winner will be randomly chosen. This means that there is a 25% chance that the player with the highest score will lose undeservedly. Feedback is the same as in the neutral treatment. Since no feedback is given about which of the two above-described scenarios has occurred, participants do not know whether they won/lost deservedly or undeservedly.

The tournament in task 3 (if chosen) will be a fair (neutral) tournament in order to both keep the treatments as similar as possible and to avoid that preferences for fair (neutral) competitive environment could play a role in task 3's decision. This treatment reproduces the features of many real world situations where individuals do not always know whether they won/lost because someone had an unfair advantage or not.

3.3 Unfair Feedback Treatment

It is analogous to the "Unfair Treatment" except that players are now told whether they won/lost deservedly or undeservedly the tournament in Task 2. Participants also receive feedback about their true rank in the "Guessing task". In this case, they do not have the possibility to update their estimated rank.

The experiment ends with a questionnaire asking questions about willingness to take risks, about whether participants think that men or women are better in the counting zeros task, demographic controls, personal and family background, sportive background, as well as questions about perception of unfairness and competitive attitudes.

4 Data Collection Procedure

In order to guarantee gender balance in our overall sample, we break down our experiment into smaller batches each with exclusively male or exclusively female participants. In order to avoid time effects, we alternate male and female batches throughout the entire data collection process. Within every batch, subjects will be randomly assigned to one of the treatments at the beginning of the experiment. Each treatment consists of three effort tasks (plus an intermediate task in which we elicit beliefs about own performances)[1] and concludes with a questionnaire. At the end of the experiment, one task will be randomly drawn to become payoff relevant. Subjects are told that there are multiple parts at the beginning of the experiment and each part is explicitly introduced and concluded using different screens. Subjects will be recruited using Prolific (see Palan & Schitter 2018). Subjects will be able to participate in the experiment only once. The interface was programmed using Qualtrics. Only subjects who are living in the US or Great Britain and that are citizens of either country will be recruited. In addition, we only invite

[1]As already mentioned in the experimental design section, the belief elicitation task consist of two measurements: one before receiving the information about the outcome of Task 2 (i.e. win or lose) and the other after having received this information. Since the belief elicitation task is a subtask of Task 2, subject will receive the bonus payment associated with it only if Task 2 is randomly drawn for payment.

subjects whose first language is English. Taken together these measures will minimize the probability of not understanding the instructions. In order to prevent non-human users from participating in the experiment, participants will have to pass a captcha-test. We will invite only subjects who have completed 10 or more surveys and whose submissions are accepted at least 95 percent of the time. Participants will receive a show up fee of £1.50 and a bonus payment which depends on the decision they will make during the experiment (see experiment instructions).

4.1 Sample restrictions

For our main analysis, we will exclude participants based on the following criteria:

1. Subjects who do not pass two fair attention checks (in line with the Prolific's Best Practice Guide).

2. Subjects who fail the first set of comprehension questions placed after Task1's instructions after two tries (they will be informed before starting the comprehension test that they will not be able to continue with the study if they do not answer correctly after two attempts)[2].

We will repeat the analysis relaxing restriction 1) in the Appendix.

5 Identification Strategy

In the following subsections we will follow the "primary outcomes section" of this preregistration.

5.1 Gender Differences in Round 3 Tournament Entry

As specified in the first point of our "primary outcomes section", we study whether willingness to enter the tournament in task 3 differs by gender. We answer this question for all three of our treatments. We do so by performing a one-sided Fisher's exact test (see section 6.1 for a power analysis).

[2]We also have a set of comprehension questions after Task 2 and Task 3 but we do not exclude participants there, since we do not want that our exclusion criteria interact with our treatments. Participants who do not pass the comprehension questions will be paid a 0.50 pounds show-up fee to reimburse them for the time they spent on the instructions. In order to avoid selection effects, they will be given this information only after failing the comprehension test twice. Yet, right at the beginning of the experiment, we make explicitly clear that they will not receive the full show up fee if they do not pass some comprehension questions.

5.2 Gender Differences in Response to Losing or Winning a Competition

We answer our first research question by investigating whether the effect of losing or winning the tournament in task 2 on the willingness to enter the tournament in task 3 differs by gender. It is important to note that, conditional on performance, winning or losing a tournament is a random event that depends on the quality of the randomly chosen competitor. By controlling for performance in a flexible way, we can therefore identify the causal effect of losing a competition. We can answer this question for all three of our treatments as a way to identify the effect of losing a competition in three types of tournaments. We do so by performing a one-sided Fisher's exact test (see section 6.2 for a power analysis).

5.3 The Effect of Unfair Competition

As specified in the second point of our "primary outcomes section", we answer our second research question by investigating whether the gender difference in response to winning or losing a competition differs by treatment. Our focus will be on comparing the "Neutral" and "Unfair" treatments. In particular, we will test whether the gender gap in willingness to enter the tournament after winning or losing a competition differs depending on the type of competition (neutral vs. unfair). We do so by using a difference in difference approach (see section 6.3 for a power analysis).

5.4 Mechanisms Driving the Effect of Unfair Competition

We investigate the causal mechanisms driving a potential effect of unfair competition. In particular, one reason why men may be less discouraged by losing an unfair competition than women, is that they may be more likely to convince themselves that their loss was undeserved (Dweck *et al.*, 1978). As specified in the third point of our "primary outcomes section", we study whether this is the case by looking at how participants update beliefs about own performances in the "Unfair" treatment. In particular, we can see whether after receiving information about the tournament's outcome (i.e. winning or losing) men and women update their beliefs differently both in terms of sign and magnitude. Moreover, we can exogenously remove the role of this causal mechanism in treatment "Unfair Feedback", where participants are explicitly told whether they won or lost deservedly or undeservedly. We can then identify the effect of this belief mechanism by testing whether the effect of unfair competition on the gender gap in tournament entry is significantly larger in the "Unfair Treatment" than in the "Unfair Feedback Treatment".

6 Power analysis

6.1 Unconditional raw gender gaps within treatments

From Buser & Yuan (2019) data we can calculate the proportion of men and women who decide to compete again in round 2 after having already competed in round 1. Since our neutral treatment replicates Buser and Yuan 2019's setting, we can use this data to calibrate the power of our tests for this treatment. Out of 93 men, 57 decide to compete again (61%), while, out of 95 women, only 44 decide to do so (46%). The gender gap is 15 percentage points. Using a one-sided exact Fisher's test with power=0.8, $\alpha = 0.05$, $p_1 = 0.61$ and $p_2 = 0.46$ we calculate that the total sample size needed to detect a minimum effect size of 15 percentage points is 288 participants. Planning to collect between 400 and 600 participants per treatment, (split equally by gender) allows us to detect such a difference. As for the unfair treatment, we can use data from our pilot experiments. We see that roughly 71% men decide to recompete again versus roughly 33% of women. Using a one-sided exact Fisher's test with power=0.8, $\alpha = 0.05$, $p_1 = 0.71$ and $p_2 = 0.33$ we calculate that the total sample size needed to detect a minimum effect size of 38 percentage points is 50 participants. Again, with our planned sample size we are able to detect such a difference.

6.2 Raw gender gaps conditional on losing within treatments

In the same line of thought, from Buser & Yuan (2019) data we can also calculate the proportion of men and women who decide to compete again in round 2 after having lost the competition in round 1. Conditional on having lost, roughly 67% of men decide to compete again while only 7% of women decide to do so. Using a one-sided exact Fisher test with power=0.8, $\alpha = 0.05$, $p_1 = 0.67$ and $p_2 = 0.07$ we calculate that the total sample size needed to detect a minimum effect size of 60 percentage points is 20 participants. With our planned sample size we are able to detect such a difference.

6.3 Difference-in-Difference Tests

These tests compare the gender gap in tournament entry across treatments. Since this hypothesis involves an interaction effect, we compute power using simulations. This requires

us to specify four parameters representing the tournament entry fractions for men and women in two treatments respectively. In order to not vary too many parameters at the same time, we fix the fraction of men and competing in one treatment, and then specify several different alternative hypotheses H_1 that increase male and decrease female willingness to compete by a similar amount. For example, for baseline tournament entry rates of 40% and 50% for women and men respectively, a 20 percentage points (pp) increase in the gender gap would lead to tournament entry rates of 30% and 60% for women and men respectively. For each of these alternative hypotheses H_1, we then simulate 1000 samples equal to the minimum (200 men and 200 women per treatment) and maximum (300 men and 300 women per treatment) sample size considered in this pre-registration. The power is then equal to the fraction of simulated samples in which we obtained a significantly larger gender gap in one treatment with a p-values smaller than or equal to 0.05.
The Table below presents power results for four different assumed effect sizes (H_1), where we use the gender gap in tournament entry in the unfair treatment of the pilot experiment as the reference gender gap. For our largest sample size, we would have good power to detect even modest changes to the gender gap (15pp), though the power is lower if we focus only on the losers (half the sample, the bottom two rows). It is also worth noting that our expected sample size is much larger than the sample sizes typically observed in this literature (see Table A2 in Buser *et al.*, 2021b for an overview of the sample sizes and effect sizes in previous work looking at differences-in-differences), and has approximately 2-3 times the number of observations observed in the laboratory experiments of Buser & Yuan (2019).

Assuming a a baseline gender gap of 50/40:

Effect Size	15pp	20pp	25pp	30pp
N=600	0.74	0.95	0.99	1
N=400	0.59	0.82	0.96	0.99
N=300	0.47	0.72	0.87	0.97
N=200	0.34	0.55	0.74	0.89

6.4 Heterogeneous treatment effects

Average treatment effects (here, the gender gap) can be estimated using a standard regression framework assuming that the treatment effect is homogeneous across individuals with different characteristics. In order to relax this assumption and explore which men and women drive the gender gap, we use machine learning methods – e.g., random forests and LASSO (for details see, e.g., Chernozhukov *et al.*, 2013; Chernozhukov *et al.*, 2017; Athey *et al.*, 2018) – that allow to consistently estimate heterogeneous treatment effects in, possibly, a large number of individual characteristics (covariates). In the case of random forests, for instance, this corresponds to run a standard regression in local neighborhoods of the covariate space within which we assume treatment effect homogeneity. In fact, random forests can be seen as (data-driven) nearest-neighbor matching estimators that detect the relevant treatment effect heterogeneities across a large number of dimensions. In particular, these algorithms estimate the relevant neighborhoods by maximizing treatment effect heterogeneity across partitions of the covariate space. In order to reduce the large dimensionality of the set of control variables and avoid the risk of overfitting, machine learning methods use bootstrapping or, more generally, random subsampling. Subsampling and randomization techniques allow machine learning algorithms to perform well also for relatively small datasets and, in particular, few hundreds of treated units as we have in our study (see e.g. simulations by Athey et al. `https://grf-labs.github.io/grf/articles/grf.html` and concrete applications for heterogeneous treatment effect estimation by Valente 2021 `https://arxiv.org/abs/2010.01105`)

References

Apicella, Coren L, Demiral, Elif E, & Mollerstrom, Johanna. 2017. No gender difference in willingness to compete when competing against self. *American Economic Review*, **107**(5), 136–40.

Athey, S., Tibshirani, J., & Wager, S. 2018. Generalized Random Forests. *Forthcoming, Annals of Statistics.*

Balafoutas, Loukas, & Sutter, Matthias. 2010 (May). *Gender, Competition and the Efficiency of Policy Intervention.* Working Papers in Economics 450. University of Gothenburg, Department of Economics.

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

Buser, T., Cappelen, A., & Tungodden, B. 2021a. Fairness and Willingness to Compete. *NHH Discussion Paper.*

Buser, T., Ranehill, E., & van Veldhuizen, R. 2021b. Gender differences in willingness to compete: The role of public observability. *Journal of Economic Psychology*, **83**.

Buser, Thomas, & Yuan, Huaiping. 2019. Do women give up competing more easily? Evidence from the lab and the Dutch math olympiad. *American Economic Journal: Applied Economics*, **11**(3), 225–52.

Buser, Thomas, Niederle, Muriel, & Oosterbeek, Hessel. 2014. Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, **129**(3), 1409–1447.

Catalyst. 2019. Women CEOs of the S&P 500. *Available at* `https://www.catalyst.org/research/women-ceos-of-the-sp-500/`.

Chernozhukov, Victor, Chetverikov, Denis, & Kato, Kengo. 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41**(6), 2786–2819.

Chernozhukov, Victor, Goldman, Matt, Semenova, Vira, & Taddy, Matt. 2017. Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels. ArXiv e-prints working paper (arXiv:1712.09988).

Croson, Rachel, & Gneezy, Uri. 2009. Gender Differences in Preferences. *Journal of Economic Literature*, **47**(2), 448–474.

Dweck, Carol S, Davidson, William, Nelson, Sharon, & Enna, Bradley. 1978. Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental psychology*, **14**(3), 268.

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

Hoogendoorn, Sander, Oosterbeek, Hessel, & Van Praag, Mirjam. 2013. The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, **59**(7), 1514–1528.

Maxwell, Sarah, Lee, Sanghyun, Anselstetter, Sabine, Comer, Lucette B, & Maxwell, Nicholas. 2009. Gender differences in the response to unfair prices: a cross-country analysis. *Journal of Consumer Marketing*, **26**(7), 508–515.

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

Palan, Stefan, & Schitter, Christian. 2018. Prolific. A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, **17**, 22–27.

Slonim, Robert, Roth, Alvin E, *et al.* 1998. Learning in high stakes ultimatum games: An experiment in the Slovak Republic. *Econometrica*, **66**, 569–596.

Solnick, Sara J. 2001. Gender differences in the ultimatum game. *Economic Inquiry*, **39**(2), 189–200.

Valente, M. 2021. Heterogeneous effects of waste pricing policies. *arXiv preprint arXiv:2010.01105*.