Never too Late: Gender Quotas in the Final Round of a Multistage Tournament

Czibor, E.; Dominguez Martinez, S.

DOI
10.1093/jleo/ewz003

Publication date
2019

Document Version
Final published version

Published in
Journal of Law, Economics, & Organization

Never too Late: Gender Quotas in the Final Round of a Multistage Tournament

Eszter Czibor
University of Chicago

Silvia Dominguez Martinez
University of Amsterdam

Affirmative action policies have been shown to induce talented women to compete in laboratory contexts. However, evidence from actual policy changes is more ambiguous. While existing laboratory experiments have exclusively analyzed gender quotas in one-shot tournaments, we focus on a setting that models real life examples, such as quotas in corporate boards, more closely: quotas implemented at the final round of a multistage elimination contest. We find that later-stage quotas increase female participation already in the first round of the tournament, showing that women are responsive to changes in the option value of continued competition. Quotas also increase high-ability women's representation among the final-stage competitors without significantly reducing entry among men, thereby promoting diversity without harming efficiency. We provide evidence for the importance of relative performance beliefs in determining the response to quotas among both genders (*JEL* C9, D9, J1, M5).

1. Introduction

Affirmative action (AA) policies guaranteeing women equal representation have been the subject of intense debates in Europe over the last decade as several countries introduced gender quotas in different areas of legislation or business.[1] Proponents of such policies argue that quotas

We are grateful to the Research Priority Area Behavioral Economics of the University of Amsterdam for their generous support. E.C. gratefully acknowledges the receipt of the Rubicon grant from the Netherlands Organisation for Scientific Research (NWO) that made her research possible. We thank seminar participants at the University of Amsterdam, the University of Chicago, the Erasmus University of Rotterdam, and the Economic Science Association World Meeting in San Diego for their insightful comments. We are particularly grateful for the helpful suggestions of Thomas Buser, Josse Delfgaauw, John List, Randolph Sloof, Joep Sonnemans, Matthias Sutter, and two anonymous referees.

1. According to the Global Database of Quotas for Women, half of the countries of the world today use some form of electoral quota for their parliament (http://www.quotaproject.org/). Rosen (2017) provides a comparative analysis of the different forms of electoral AA policies in developed as well as developing countries. In business, Norway pioneered by obliging companies to ensure that at least 40% of board members are female by January 2008 (Ahern and Dittmar 2012; Matsa and Miller 2013). Germany has very recently passed a

The Journal of Law, Economics, and Organization, Vol. 35, No. 2
doi:10.1093/jleo/ewz003
Advance Access published April 22, 2019

do not imply a trade-off between equity and efficiency: rather, they help correct inefficiencies stemming from the fact that talented women do not get hired or promoted otherwise. This may happen either because of discrimination in the selection process (e.g., Goldin and Rouse 2000; Reuben et al. 2014; Beaurain and Masclet 2016), or simply because high-ability women choose not to enter the contests in the first place. Due to their lower levels of confidence, higher levels of risk aversion and greater distaste for (mixed-sex) competition, women have been shown to shy away from tournaments they could win (e.g., Niederle and Vesterlund 2007, 2011; Croson and Gneezy 2009; Azmat and Petrongolo 2014).[2] AA policies may thus improve the quality of the applicant pool if they induce high ability women to compete without discouraging men.

Recent experiments present evidence in support of this claim both from the laboratory (Balafoutas and Sutter 2012; Niederle et al. 2013) and from the field (Ibanez and Riener 2018).[3] However, results regarding the impact of real-life AA policies on female tournament entry are more ambiguous: while gender quotas in village councils raised career aspiration and educational attainment for girls (Beaman et al. 2012), the quota for Norwegian company boards left younger women's plans regarding business school enrollment or fertility largely unaffected (Bertrand et al. 2019), bringing into question the assumption that quotas attract talented women to competition. The above-mentioned experimental studies on gender quotas differ from the Norwegian case along a key dimension: while the former all focus on one-shot tournaments, the latter can be modeled as a *multistage elimination tournament* (Rosen 1986), where the AA policy only applies in the final round. As a result, women have to enter and win a sequence of tournaments without AA in order to enjoy the benefits of the quota at the last stage.

It is unclear whether results obtained in one-shot settings carry over to sequential elimination contests. Altmann et al. (2012), analyzing effort provision in multistage elimination tournaments, find contestants to be forward-looking: participants consider the option value of future promotion possibilities when deciding how hard to work in the first round.

law requiring at least 30% of non-executive members at large companies to be female by 2016. Other countries such as The Netherlands have less strict "comply or explain" policies in place for company boards (non-complying firms have to explain in their annual reports why they failed to increase female representation).

2. Gender differences in competitive preferences have been also linked to the gender wage gap, see for example, Booth (2009) and Grund (2015).

3. A few papers rely on natural experiments to study the impact of gender quotas on the quality of "winners" (broadly defined to encompass the group of people hired/promoted/elected for the particular position in question) without explicitly considering the change in the group of *applicants*, see for example, Ahern and Dittmar (2012) and Baltrunaite et al. (2014). Other papers abstract away from possible changes in group composition and focus on the causal impact of quotas on effort and performance (Schotter and Weigelt 1992; Bracha et al. 2013; Calsamiglia et al. 2013). In an important paper, Beaman et al. (2009) measure the long-run impact of AA on discriminatory attitudes among voters.

Importantly, the authors also find that contestants provide excess effort in the first round (both compared with Nash predictions and to subjects' behavior in a treatment that involves a strategically equivalent one-shot tournament). Altmann et al. (2012) attribute this higher effort exertion to an overestimation of the option value as well as to a non-pecuniary taste for competing. Given the gender differences in preferences and confidence discussed above, we expect both of these channels to apply more to male than to female contestants.[4] As a result, we predict that the gender gap in the willingness to enter one-shot tournaments also appears in multistage elimination contests where preferences and beliefs influence not only participants' expectations of the direct returns from competing in the given round of the tournament, but also the perceived value of future competition. This may be particularly true for high ability participants who can realistically expect to continue competing beyond the first round.

Predictions for the impact of a final-round gender quota in a multistage contest are more ambiguous. Quotas in a later stage likely affect the entry decision already in the first round by changing the option value of competing in further stages. For women (men), AA policies increase (decrease) this option value by increasing (decreasing) the probability of winning the last-stage tournament. Besides changing the objective probability of winning, quotas essentially create a single-sex tournament for women in the final round, making the tournament more attractive for women (Gupta et al. 2011; Grosse et al. 2014). However, the change in the option value of continued competition depends on participants' beliefs about their relative performance. For contestants with low expected performance in the first round (and thus small expected probability of continuation), quotas in later stages are practically irrelevant. If women are pessimistic about their ability to win the competition in the first stage, they may shy away from the tournament early on, and thus miss out on future benefits of the quota. On the other hand, participants who expect to be top performers believe they will win the second stage anyway, irrespective of the quotas. Consequently, (over-)confident men may not respond negatively to gender quotas. We thus anticipate the largest first-stage response to later-stage quotas among those with intermediate rank guesses, both men and women.

To test these hypotheses, we conduct a real-effort laboratory experiment comparing the willingness to enter a two-stage elimination contest with or without AA policy in the second round. Specifically, we test the impact of a quota that reserves half of the winner positions in the second stage for women. Importantly, we design the experiment such that quotas leave the first-stage probability of winning unchanged, so any difference in the propensity to compete in the early stage must result from changes in the option value of future competition possibilities. Moreover, we keep the

4. Altmann et al. (2012) did not report the gender composition of their sample, nor discuss potential heterogeneity of behavior by gender.

winners' prizes constant across the rounds to estimate the impact of quotas separately from the effect of higher rewards on the gender gap in entry (Petrie and Segal 2014; Ifcher and Zarghamee 2015). Besides entry rates, we also analyze the gender composition and performance of the pool of winners to assess whether quotas result in a more equal representation without a loss in efficiency. Finally, we test whether the AA policy has a "spillover effect" on altruistic behavior by eliciting choice in a subsequent dictator game. In order to keep our design simple and tractable, in our study we abstract away from human capital acquisition (Cotton et al. 2014; Stark and Hyll 2014).

Our results largely confirm the predictions outlined above. In the absence of AA policies we find a significant gender gap in the propensity to compete among high ability participants in both rounds of the elimination contest, resulting from high performing women shying away from tournaments they could win. We do not observe a gender gap in entry rates among low ability subjects. Second-stage quotas attract high performing women to the tournament already in the first stage, and ensure that female entrants continue to compete in the second stage. The effect of quotas is concentrated among those with intermediate relative performance beliefs. Contrary to our hypotheses, we observe no significant reduction in men's propensity to compete in response to the AA policy, a finding that is largely attributable to men's overconfidence. As a result, we find no negative effect of the AA policy on efficiency: quotas do not lead to lower performance overall, nor do they cause "reverse discrimination" (there is only one instance in our data when a men with a higher score is passed over to promote a woman in order to fulfill the quota requirements). We do not detect any negative impact of the quotas on subsequent altruistic choices, if anything, male dictators are more generous in the treatment with AA.[5]

In sum, we find that final-stage quotas increase female representation in multistage elimination tournaments without discouraging men, and therefore ensure gender diversity while maintaining the quality of winners. We thus show that results obtained in one-shot tournaments (Balafoutas and Sutter 2012; Niederle et al. 2013) replicate in a setting that resembles real-life examples of AA policies more closely, such as quotas imposed for board of directors or high-level legislative bodies. Importantly, quotas in the final round seem to address both the issues of initial self-selection and later retention: they encourage women to enter the first round, and to keep competing in the subsequent rounds. Our results are in line with the

5. This result is in line with Banerjee et al. (2016)'s findings that quotas do not increase unethical behavior or spite toward members of the advantaged group. Relatedly, Maggian and Montinari (2017) find no negative spillovers of gender quotas on subsequent unethical behavior either among men or among women. Balafoutas and Sutter (2012) find no negative impact of AA policies on subsequent cooperation. Similarly, Koelle (2016) finds that quotas do not discourage people from forming teams in a later stage, nor do they harm performance in teams.

findings of the only other study of gender quotas in a multistage elimination contest that we are aware of: in an experiment conducted concurrently to our own, Maggian et al. (2017) also find that second-stage quotas eliminate the gender gap in entry both in the first and the second round of the tournament. Our study complements their findings by exploring potential mechanisms: we highlight the importance of beliefs in driving the response to quotas both among male and female participants. We also examine the impact of AA policies more broadly, by including among our outcomes the behavior in a subsequent dictator game.

We note that last-stage quotas encompass by design some form of "quality control": under this mechanism, only those who compete and win in previous rounds are eligible for the reserved winner positions in the final round, ensuring that the lowest performing women do not qualify for promotion under the quotas. This desirable feature of later-stage quotas may make them preferred to policies implemented in earlier stages, and could increase popular support for their introduction.

The remainder of this paper is structured as follows. In Section 2, we present the context and design of the experiment. Section 3 contains an overview of our data. Section 4 presents the result, while Section 5 discusses the external validity of our findings. Section 6 concludes.

2. Context and Design

Our experiment was conducted at the CREED laboratory of the University of Amsterdam. We ran 12 experimental sessions between January and September 2016. Due to differences in show-up rates, session sizes varied between 16 and 30 participants.[6] In total, 274 subjects participated in the experiment, half of them female. Sessions lasted approximately 1 h, and participants earned on average €17.4, including an €8 show-up fee. The experiment was conducted in English and programmed using zTree (Fischbacher 2007). Instructions are included in the Appendix.

Our experimental design extends the frameworks of Niederle et al. (2013) (NSV2013) and Balafoutas and Sutter (2012) (BS2012) who compare the willingness to compete in standard tournaments and under different AA policies. While NSV2013 and BS2012 analyze one-shot tournaments, we focus on a two-stage elimination contest, and study the impact of gender quotas implemented in the second stage. As such, our design models AA policies that target women later in their careers and higher in the hierarchy, such as the increasingly common but widely debated gender quotas for corporate boards. We use a between-subject design such that participants are assigned either to the Control or the AA treatment.[7]

6. An overview of the number of subjects and the share of female participants per session is presented in Table A1 in Appendix A.

7. In this sense, our experiment also serves as a robustness check for the results of NSV2013 who used a within-subject design, where potential confounds such as sensitization to changes, learning effects, or other contextual factors may be of concern (Charness et al.

Our experiment consists of three "main" rounds assessing participants' performance and choices in competitive environments, followed by additional tasks eliciting risk preferences, confidence, and altruism. One of the three rounds is randomly chosen at the end of the experiment to be paid. Figure 1 presents an overview of the rounds.

In each of the main three rounds, participants have 4 min to work on the same real-effort task under varying compensation schemes: as in BS2012 and NSV2013, the task involves calculating the correct sum of sets of five 2-digit numbers (e.g., $27 + 31 + 85 + 11 + 76 = ?$). At the end of each round, participants learn the number of problems they solved correctly, but receive no relative performance feedback.

Round 0—Tournament. After basic demographic questions and an unpaid practice round, in Round 0 all participants are assigned to a tournament scheme where the top 50% of performers in the session are selected as winners and earn €1 per correctly solved problem, while others receive no compensation (ties are broken randomly).[8] This round allows participants to experience competition and provides us with a measure of their "baseline" task performance in a competitive environment. Participants are not informed of whether they are among the winners until the end of the experiment.

Rounds 1 and 2—Two-stage elimination contest. In Round 1 of the experiment, before starting the new set of addition tasks, participants are asked to select the remuneration scheme that will determine their earnings from this round. They have a choice between a piece rate scheme that pays 50 cents per correctly solved problem or a tournament scheme where winners earn €1 per correct answer and others receive no payment. Winners of the tournament are determined by comparing participants' Round 1 performance to the Round 0 performance of all the other subjects in their session: tournament entrants who score higher than all non-winners did in the previous round are selected to be winners. Letting participants compete against others' past performance makes our results from this stage comparable to existing studies on the gender gap in competition entry (including NSV2013 and BS2012) who use this design to measure competitive preferences independent of altruism or beliefs about others' selection decisions (Niederle and Vesterlund 2007).

2012). To avoid contamination, all participants in a given session were subject to the same treatment condition, and we ran six sessions per treatment.

8. Throughout the experiment, we let participants compete against all others in their session rather than randomly assign them to smaller groups to compete against, as BS2012 and NSV2013 did. This choice reduces the noise in tournament outcomes stemming from random differences in the quality of one's opponents (that would otherwise be especially pronounced in the second round that by definition has fewer entrants). It also helps avoid a situation where in the AA condition reserved seats go unfilled in one group despite there being more female entrants than reserved seats in another group in the same session.

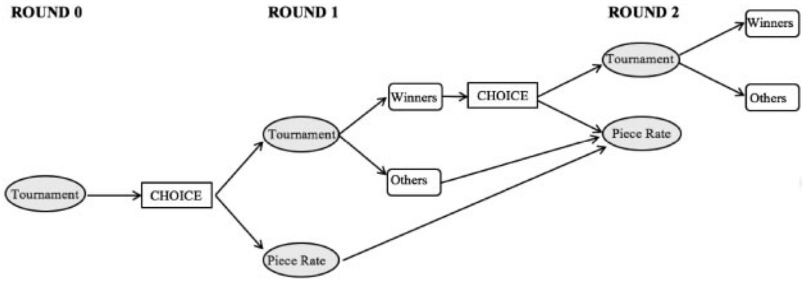


Figure 1. Overview of the Experiment.

Crucially, when making their remuneration choice in Round 1, participants are informed that their decision and task performance has consequences for their options in Round 2. In particular, only winners of the Round 1 tournament are given a choice again between a piece rate and a competitive scheme in Round 2 (with the same piece rate and prizes as in Round 1). Those who select the piece rate option in Round 1, or enter but do not win the tournament in Round 1, are automatically assigned to a piece rate scheme in Round 2. Rounds 1 and 2 thus correspond to the first and second stages of a multistage elimination contest. Since participants only learn at the end of the experiment whether or not they were winners, we use a strategy method in Round 2 to elicit the choice between piece rate and tournament among all those who chose to compete in Round 1. In order to ensure that they understand the rules, participants answer control questions about the design before making their choice in Round 1.

The rules determining the winners in Round 2 tournament differ between the two treatments. In the Control treatment, the 50% of tournament entrants with the highest Round 2 scores are chosen as winners.[9] In the AA treatment, half of the winner positions are reserved for female entrants. Specifically, participants receive the following information on the procedure we use to select winners:

> First, we rank all participants who entered the tournament within their gender group based on their Round 2 performance: females compared with the group of females, males compared with the group of males. (When there are ties in the ranking, they are broken at random.)

- If more female participants enter the tournament than winner positions reserved for women, then the females with the highest rank are placed in the reserved positions. The other female

9. Ties are broken randomly; in case an odd number of participants enter the tournament, the number of winners is randomly rounded up or down to the nearest integer. “Tournament entrants” refers to participants who selected the tournament option for Round 2 and who were winners in Round 1.

entrants are grouped together with all the male entrants, and a new ranking is determined within this group. The highest ranked participants from this group fill the non-reserved winner positions.

- If at most as many female participants enter the tournament as the number of winner positions reserved for women, then all females who entered are selected as winners. All the remaining winner positions (i.e., reserved but not filled and non-reserved) are filled with the highest ranked male participants.

Table 1 summarizes how tournament winners are determined in each round in the two treatments.

The following three features of our design are worth emphasizing. First, since participants do not receive any information about the AA condition until the beginning of Round 1, the treatments do not affect the "baseline" performance measure we elicit in Round 0. Second, since participants in the Round 1 tournament compete against the past performance of all others in their session, the AA treatment does not influence their chances of winning in that round. Any difference we observe in Round 1 tournament entry rates between the two treatments is thus attributable to the *indirect* impact of future quotas, brought about by changes in the option value of continued competition resulting from the later-round quotas. In Round 2, however, entrants are measured against other entrants' current performance, thus beliefs about the composition of the pool of opponents could influence the decision to compete. This aspect is very relevant, since the multistage elimination contest setting ensures by design that final-round competitors are positively selected on performance among the entrants (remember, only Round 1 winners are allowed to compete in Round 2). For this reason, we believe that the meaningful entry decisions to study in the final round is the one that is affected by beliefs about the quality of opponents. Competing against other successful entrants' current performance thus allows us to capture the *direct* effect of quotas on entry rates in the round where they are applied, stemming both from changes in the probability of winning (keeping competitors constant) and from changes in expectations around the composition of the competitor pool.[10] Third, we decided to keep winners' prizes constant across the two stages of the elimination contest despite the fact that the original model of sequential elimination tournaments involved higher Round 2 prizes (Rosen 1986). This choice was motivated by findings from Ifcher and Zarghamee (2015) and Petrie and Segal (2014) who demonstrated the crucial role that the size of the reward plays in determining the gender gap in competition entry.

10. Note that the number of Round 2 winners in our study is endogenously determined: the number of prizes is increasing with the number of entrants, such that it is only the quality, not the number, of opponents that determines one's chances of winning. We discuss the implications of this design choice in Section 5.

Table 1. Determining Tournament Winners

	Control	AA
Round 0	Top 50% in session	Top 50% in session
Round 1	Those who score higher than non-winners did in Round 0	Those who score higher than non-winners did in Round 0
Round 2	Top 50% of entrants in Round 2	Top 50% of entrants in Round 2, *half of winner positions reserved for women*

Our goal was to isolate the impact of final-round gender quotas from any potentially confounding prize effect, so rather than trying to model the interaction between higher prizes and quotas, we decided to keep final-round prizes the same as in earlier rounds.

After the three rounds focusing on competitive choices, we measure participants' risk preferences through an incentivized choice between risky gambles (Eckel and Grossman 2002) as well as an unincentivized self-assessment question (Dohmen et al. 2011).[11] Participants' relative performance beliefs are then elicited in various scenarios. In particular, they are asked to report their guessed ranks (on a scale of 1–4, where 1 corresponds to the best 25% in the session and 4 corresponds to the worst 25% in the session) in terms of task performance (i) in Round 0, relative to all other participants in the session; (ii) in Round 1, relative to others' Round 0 performance; (iii) in Round 2, relative to all others who have entered the tournament in that part; (iv) in Round 2, relative to all other participants of the same sex who entered the competition in that part. We also ask participants for the share of female tournament entrants they expect in Round 2. These guesses are incentivized: participants are informed that one of the above subquestions will be randomly chosen at the end of the experiment, and a correct answer is rewarded with €1. Finally, we use a dictator game to measure social preferences: using a strategy method, we ask participants what share of a €2 endowment they would like to send to the recipient in case they are assigned the role of sender.[12] The experiment concludes with open questions asking participants to explain their choices in Rounds 1 and 2.

11. Participants are first offered a choice between five investment opportunities that each have a 50% chance of succeeding but differ in the payoffs associated with being unsuccessful/successful, ranging from €0.8–€0.8 (Project A) to €0–€2.4 (Project E), see Figure A1 in Appendix A. They are then asked the following question: *How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please choose a value on the scale below, where the value 0 means "unwilling to take risks" and the value 10 means "fully prepared to take risks"* (Dohmen et al. 2011).

12. Participants can select the amount they wish to send from a list of 11 options spanning the range €0–€2 in 20 cents increments. Due to a small coding error, the list of options indicates that sending €0.2 results in keeping €2 instead of €1.8, see Figure A2 in Appendix

3. Data

Our sample consists of 274 participants, 137 of whom are female. Participants are on average 22.3 years old. Overall, less than 60% of our subjects are Dutch, and approximately 60% follow an Economics and Business study specialization. The mean number of addition tasks solved correctly in Round 0 is 9.7 (with a standard deviation of 4.2), while the median is 9 (see Figure A3 in Appendix A). Task performance improves over time (mean scores are 10.6 and 11.0 in Rounds 1 and 2, respectively), possibly due to learning effects.

Table 2 compares participants assigned to the Control and the AA conditions along several demographic characteristics and experimental measures. Comparisons are carried out within gender subsamples; the first four columns of the table refer to male, the last four columns to female participants. While the subsample of women is balanced between the treatments along all dimensions, we find significant differences between the two treatment groups among men. In particular, among male participants the share of Dutch citizens is lower and the amount sent to the receiver in the dictator game is higher in the AA treatment. More concerning is that men in the AA treatment score significantly lower on the task in Round 0 than those in the Control, suggesting that there is a difference in "baseline" ability between the two groups. As a result, while there is no gender difference in performance in the AA condition, as is typically the case with the addition task (see e.g., Niederle and Vesterlund 2007), men outperform women in the Control group. These differences in scores carry over to subsequent rounds and are highly statistically significant. We would like to emphasize that this issue does not bias between-treatment comparison among women as they received no performance feedback during the experiment, and their incentivized relative performance beliefs do not reflect this gap. Admittedly, it slightly complicates the analysis of the subsample of male participants, of gender differences in the Control condition, and the study of efficiency. Throughout Section 4 we supplement raw comparisons by treatment or gender with results from a regression framework where we control for baseline task scores in order to correct for differences caused by performance disparities unrelated to our treatment.

Table 2 also allows for gender comparisons along demographic characteristics and preferences. We observe that women in our sample are less likely than men to be Dutch and tend to have higher high school mathematics grades (differences that are significant on the 1% and 5% level, respectively). Based on their choices between risky gambles, we find women to be more risk averse than men, confirming results by Charness and Gneezy (2012) and Eckel and Grossman (2008).

A. To avoid deception, actual earnings were calculated in adherence to the amounts announced.

Table 2. Comparison of Means between Treatment Groups, by Gender

	Men ($N=137$)				Women ($N=137$)			
	Control	AA	Difference	*p*-value	Control	AA	Difference	*p*-value
Age	22.84	22.10	0.74	0.17	22.28	21.86	0.43	0.52
Dutch	0.78	0.57	0.20	0.01	0.54	0.47	0.07	0.44
Econ	0.57	0.67	−0.10	0.21	0.54	0.63	−0.09	0.28
Math grade	7.21	7.41	−0.20	0.35	7.72	7.71	0.01	0.98
Score Round 0	11.37	9.26	2.12	0.00	9.18	9.23	−0.05	0.94
Score Round 1	12.39	10.23	2.16	0.01	9.76	9.89	−0.12	0.85
Score Round 2	12.46	10.81	1.65	0.04	10.22	10.56	−0.33	0.61
Guessed rank Round 0	1.90	1.94	−0.05	0.76	2.03	2.14	−0.11	0.45
Risk taking (incentivized)	3.96	3.74	0.21	0.36	2.93	3.16	−0.23	0.32
Risk taking (self-assessed)	6.63	6.39	0.24	0.51	5.39	5.61	−0.23	0.60
Dictator giving	17.61	28.57	−10.96	0.00	27.46	25.14	2.32	0.52

The table displays comparisons of means between the Control and the AA treatment conditions, by gender. Columns (2)–(5) refer to the subsample of male, Columns (6)–(9) to female participants. The *p*-value of the differences in means is calculated by *t*-tests with unequal variances.

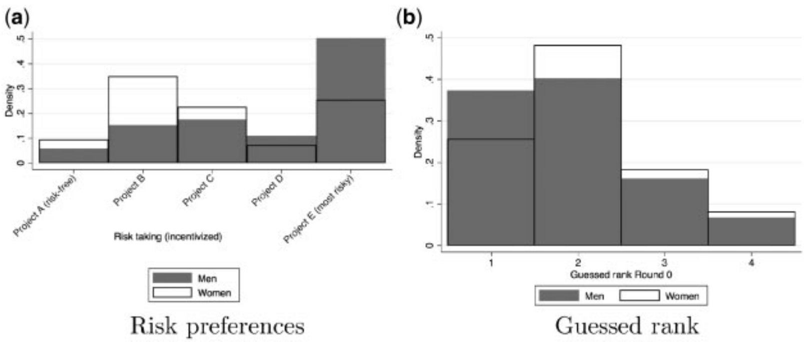


Figure 2. Gender Differences in Preferences and Relative Performance Beliefs.

As seen in Figure 2a, men are particularly more likely than women to select the riskiest investment option that has the highest expected value but entails the possibility of earning nothing, that is, Project E. Participants' self-assessments confirm the existence of a gender gap in risk preferences, as in Dohmen et al. (2011). There is, however, no clear gender difference in relative confidence: Figure 2b shows that guessed ranks in terms of Round 0 performance are very similar among men and women. The modal answer in both groups is a rank guess of 2 (corresponding to beliefs of being in the top 25–50% of the performance distribution), while 1, the top rank is the second most frequently selected. A Kolmogorov–Smirnov test for the equality of distributions yields a p-value of 0.308 for relative performance beliefs. This finding is in contrast with the empirical stylized fact that women are less confident (e.g., Croson and Gneezy 2009) and is all the more surprising given that in our setting men in the Control group actually outperform women.[13]

4. Results

In the following we analyze participants' tournament entry choices to test the predictions derived in Section 1. First, we study whether the gender gap in selection into competition observed in one-shot tournaments (Niederle and Vesterlund 2007) carries over to a multistage elimination contest and results in high-ability women missing from among the contestants in the second round. Second, we assess how second-stage gender quotas affect tournament entry rates among male and female participants in each round of the two-stage elimination contest. Third, we explore what mechanisms explain the reaction to AA; in particular, we test whether relative performance beliefs are important predictors of the response. Finally, we study the consequences of gender quotas for female representation, for efficiency and for other-regarding behavior.

4.1 The Impact of AA on the Selection into Competition

We begin our analysis by a simple comparison of tournament entry rates in Rounds 1 and 2 between the two treatments. In the following discussion, unless otherwise specified, we report results from two-sided t-tests comparing the share of entrants by gender and/or treatment group.

4.1.1 Full Sample. Panel (a) of Figure 3 depicts the share of participants who select the tournament option in each round, by treatment and gender. In the Control condition, we observe moderately high entry rates in the first round: 61.2% of women and 70.1% of men select the competitive option (note that if performance in Round 1 is exactly as in Round 2, rules of the contest are such that half of the participants have higher expected earnings from the tournament than the piece rate).[14] The gender gap in first-round entry rates in the Control group is small and insignificant (a difference of 8.9 percentage points, with a p-value of 0.278).[15] Disparities

13. Figure A4a in Appendix A depicts the distribution of risk preferences based on the self-reported, non-incentivized measure, while Figure A4b shows participants' self-assessments of their relative performance in Round 1 compared with other participants' Round 0 scores, proving that the above results are not sensitive to the particular mode of elicitation.

14. As a comparison, in the seminal paper of Niederle and Vesterlund (2007), 35% of female and 73% of male participants chose tournament over piece rate, despite only 30% having strictly higher expected earnings from the tournament option.

15. This finding, while surprising, is in line with the results of Buser and Yuan (2016) who study the dynamics of competitiveness over the same time period recruiting participants from the same potential subject pool as ours (through the CREED lab's mailing list), and find no significant gender gap in the initial propensity to compete in their Feedback experiment. To

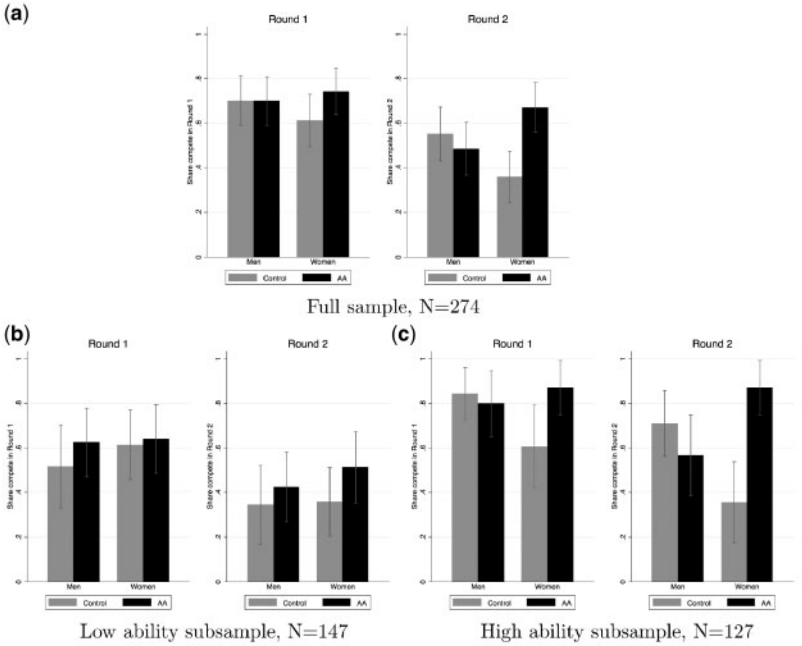


Figure 3. Share Choosing to Compete in Rounds 1 and 2 (Over All Participants in their Treatment).

are more pronounced in Round 2: women in the Control condition are much less likely than men to choose to compete in this round (35.8% versus 55.2%, the p-value of the difference 0.024). Overall, the share of participants selecting the tournament option is slightly higher in the AA treatment, where 72.1% of all participants choose to compete in the first stage. The increase is driven by women's choices (74.3% of female participants assigned to the AA condition choose to enter Round 1 tournament). The difference between treatments in Round 1 female tournament entry rates is 13.1 percentage points (with a p-value of 0.102). We observe a sizable and highly significant difference of more than 30 percentage points in women's willingness to enter Round 2 tournament between the two treatments. This higher rate of entry in the AA condition is largely attributable to more women choosing to continue the competition in the second round conditional on having entered the two-stage contest. Men are by and large unaffected by gender quotas: the decrease in male tournament entry rates in response to the AA condition is small and insignificant in both rounds.

avoid contamination through multiple exposure to similar treatments, in the recruitment process for the winter and spring sessions we automatically excluded all subjects who participated in Buser and Yuan (2016)'s experiment.

These raw comparisons, however, are not necessarily informative given the ability dispersion in our sample. First, remember that gender quotas in our setting only affect winning probabilities in the second stage of the elimination contest. Consequently, they only change the incentives to compete in the first stage for participants who believe they have a non-negligible chance of winning in that stage and thus are able to compete again in the subsequent stage. To phrase it differently, we expect that responsiveness to second-stage tournament rules increases with one's probability of continuation. We thus predict a stronger response to the AA treatment among high-ability participants. Second, as mentioned in Section 3, due to imperfect randomization, men in the Control condition have on average higher Round 0 scores than men in the AA treatment (and than women in the Control group). Some of the raw differences we observed in the full sample could thus potentially be attributed to performance disparities rather than true gender differences in competitive preferences or a response to the quotas. Finally, the main motivation for this study is to address the concern that top performing women shy away from competition, and to test whether second-stage quotas in a multistage elimination contest can induce them to compete more without discouraging talented men. This consideration also warrants a special focus on high ability participants.

4.1.2 Heterogeneity by Ability. We continue our analysis by repeating the above comparisons separately by ability categories. Specifically, we classify participants as high (low) ability if their Round 0 score is higher than (lower or equal to) the median score of 9. This split results in 68 men and 59 women labeled as "high ability" and 69 men and 78 women labeled as "low ability."[16] Assuming that performance in Round 1 is exactly as in Round 0, all participants classified as high ability can expect to be winners in Round 1 with certainty. Choosing the tournament option in Round 1 thus yields higher expected earnings for them than the piece rate scheme even without taking into consideration the option value of Round 2 competition. In Panels (b) and (c) of Figure 3 we compare tournament entry rates separately for the subsamples of low and high ability participants. As predicted, we find large differences in behavior between the two groups. Unsurprisingly, overall propensity to compete is greater among participants with higher Round 0 scores. In the low ability subsample there are no significant gender differences in tournament entry rates in the Control group: if anything, female participants compete slightly more than males (61.5% of women versus 51.7% of men in Round 1 and 35.9% of women versus 34.5% of men in Round 2 prefer the tournament over the piece rate

16. Round 0 performance is a good predictor of both Round 1 and Round 2 scores, and the strength of the association between baseline and future performance does not differ by treatments or by gender. Our results remain qualitatively similar if we also classify participants with a round 0 score of 9 as "high ability."

option). There are no significant differences in the willingness to compete between treatments among either male or female participants in this subsample, although tournament entry rates are slightly higher in the AA than in the Control condition in both rounds among women and—contrary to our predictions—also among men.

Among high ability participants, we find a clear gender gap in the propensity to compete in the Control treatment: in this subsample, 84.2% of men versus 60.7% of women enter the tournament in Round 1, a difference that is significant at the 5% level (p-value = 0.031). The gap is even more pronounced in Round 2 where 71.1% of high ability men and 35.7% of high ability women choose to compete in the Control condition (p-value: 0.004). The result that low ability women self-select into competition at similar rates as men while there is a gender gap in the propensity to compete among high ability participants is in line with the findings of Niederle and Vesterlund (2007) and Niederle et al. (2013) who also observe larger gender differences in tournament entry rates among participants with a high probability of winning.

In the high ability subsample, AA measures affect female behavior in both rounds: even though quotas only regulate the share of female winners in the second stage, already in Round 1 we observe significantly higher tournament entry rates among high ability women in the AA treatment than in the Control (87.1% versus 60.7%, the p-value of the difference is 0.020). The difference between treatments in the propensity to compete is even starker in Round 2: while in the Control condition only about a third of the high ability women compete, in the AA treatment this share is 87.1% (i.e., all women in this category who entered the two-stage tournament actually wish to continue competing in the second round).[17] While male tournament entry rates are lower in the AA than in the Control condition (by 4.2 and 14.4 percentage points in the first and second round, respectively), these differences are not significant at any conventional level, and rather than reflecting genuine treatment effects, might result from pre-existing ability differences between men in the two conditions (see our discussion on the imperfect balance between the two treatment groups in Section 3), a possibility we explore in the next subsection.

4.1.3 Controlling for Baseline Score. In the following, we show in a regression framework that our results are robust to controlling for baseline ability (i.e., Round 0 scores). This is a crucial step given the baseline score differences between male participants in the Control and the AA

17. Given that we compare four subgroups (men/women * high/low ability) along two different outcomes (tournament entry in Rounds 1 and 2), we need to correct for multiple hypothesis testing. Using the procedure outlined in List et al. (2016) we find that the difference among high ability women in entry rates in Round 1 is less precisely estimated (p-value: 0.117), while in Round 2 it remains highly significant (p-value: 0.003).

Table 3. The Impact of AA on the Willingness to Compete

	Full sample		High ability subsample	
	Round 1 (1)	Round 2 (2)	Round 1 (3)	Round 2 (4)
Female	−0.020	−0.116	−0.180*	−0.281**
	(0.078)	(0.082)	(0.100)	(0.111)
AA	0.066	0.009	0.014	−0.070
	(0.077)	(0.081)	(0.099)	(0.109)
AA * Female	0.064	0.302***	0.241*	0.572***
	(0.109)	(0.114)	(0.143)	(0.158)
Score Round 0	0.032***	0.036***	0.029**	0.038***
	(0.007)	(0.007)	(0.011)	(0.012)
Constant	0.341***	0.145	0.422**	0.156
	(0.093)	(0.097)	(0.173)	(0.191)
N	274	274	127	127
R^2	0.089	0.137	0.109	0.208
p(AA+AA * Female=0)	0.091	0.000	0.014	0.000

The table displays OLS estimated coefficients from linear probability models. The dependent variable is a binary indicator for selecting the tournament option (column headers indicate the relevant round). Columns (1)–(2) present estimates for the full sample, while the analysis presented in Columns (3)–(4) is restricted to the subsample of high ability participants. AA denotes the AA treatment, AA * Female is the interaction term. p(AA+AA * Female=0) presents p-values from F-tests testing whether the sum of the estimated coefficients differs from zero. Standard errors in parentheses

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

treatment. Throughout the paper, we report OLS estimates of linear probability models.[18]

We present our findings in Table 3. The main take-away is that our results remain by and large unchanged once we control for baseline score differences. In the Control group, the gender differences in tournament entry rates are not significantly different from zero in the full sample: the point estimates for the gap are 2 and 11.6 percentage points in Rounds 1 and 2, with standard errors of 0.078 and 0.082, respectively. The regression results confirm that the gap in competition entry is larger among high

18. We chose to use OLS for the ease of interpretation of the estimated coefficients associated with the interaction terms. Reassuringly, in our data there are only six instances when predicted probabilities lie outside the unit interval, suggesting that unboundedness is unlikely to seriously threaten the consistency and unbiasedness of our estimates. For a discussion on using OLS on linear probability models please refer to, for example, Horrace and Oaxaca (2006). Our results are similar if we estimate probit models instead. Given that we provide no relative performance feedback during the experiment and participants never learn others' choices, assuming uncorrelated errors on the session level is a good approximation. However, we checked that our results carry through when we cluster standard errors on the session level (given the low number of clusters in our data we use Cameron et al. [2008]'s wild cluster bootstrap-t procedure). Including session-level controls (the number of participants and the gender composition of the sessions [Fréchette, 2012]) does not change the size of the estimated coefficients but makes the estimates less precise. All calculations are available from the authors upon request.

ability participants in the Control group, especially in the second stage (an estimated gap of 18 and 28.1 percentage points in Rounds 1 and 2, significantly different from zero at the 10% and 5% level, respectively).

Our estimates confirm that the AA treatment increases female competition entry, particularly in the second round (when quotas are actually in place), and they also suggest that quotas did not discourage men from competing. Focusing on the full sample, the point estimates associated with the AA treatment (measuring male participants' response) in the two rounds are 0.066 (SE = 0.077) and 0.009 (SE = 0.081), while the interaction effect between female and AA (measuring women's differential response to the quotas compared with men) is 0.064 (SE = 0.109) and 0.302 (SE = 0.007), such that the total effect of the AA treatment on female participants is significantly different from zero at the 10% and 1% level in Rounds 1 and 2, respectively (see p-values from F-tests reported in the last row of Table 3). Results in Columns (3) and (4) further confirm that the effect on female tournament participation is greatest among high ability women (who have realistic chances of winning the tournament). The estimated treatment effect among high ability men is small in size and not statistically significant (controlling for baseline ability, the point estimates suggest a 1.4 percentage points increase in entry rates in the first round and a 7 percentage points decrease in the second round as a result of the quotas).[19]

We continue by testing to what extent the gender gap in tournament entry rates between high ability men and women is explained by factors other than competitiveness. We re-estimate the model presented in Table 3 and include measures of risk aversion and relative confidence as additional controls. Results are shown in Table A3 in Appendix A. While both risk preferences and guessed rank are important predictors of the choice to compete, we find that controlling for them hardly affects the size of the estimated gender gap in the Control condition. This suggests that these variables only explain a small portion of the gender difference in the willingness to compete. The estimated impact of the AA treatment remains mostly unchanged when we control for risk aversion and confidence, suggesting that the quotas do not operate through their impact on participants' risk preferences and relative performance beliefs.[20]

Finally, we argue that our results are robust to the small differences between sessions in the share of female participants resulting from imperfect show-up. Figure A5 in Appendix A shows for each session the share of female and male participants choosing to enter the second round of the tournament plotted against the share of women in the given session, by

19. Results from the subsample of low ability participants are presented in Table A2 in Appendix A.

20. Controlling for demographic characteristics such as age, citizenship, college major, and high school mathematics grades leaves our results qualitatively unchanged. Calculations are available from the authors on request.

treatment. We find no clear relationship between the share of female participants and the response to AA.[21] This is unsurprising given that participants were never told the exact share of men/women in their session, and deviations from gender balance were small enough to arguably go unnoticed for the casual observer. We summarize our results from this subsection below:

Result 1: In the Control condition, we replicate in a multistage elimination contest the empirical stylized fact obtained from one-shot tournaments that high ability women shy away from competitions they could win. We estimate that only a small part of the gender gap in tournament entry rates is attributable to gender differences in risk preferences or confidence. We find no gender gap in the propensity to compete among low ability participants.

Result 2: AA increases tournament entry rates among high-ability women in both the first and the second round of the multistage contest, while it does not significantly affect the willingness to compete among men or low-ability women.

4.2 Exploring Potential Mechanisms

We dedicate this subsection to studying the mechanisms underlying the response to gender quotas. The evidence presented in the previous subsection suggests that quotas affect first-stage entry rates through changing the option value of continued competition: it is the group of high ability women (who have good chances of making it to the second round) who seem most responsive to the final-stage quotas, and the majority of those women who enter the multistage contest in the AA condition then continue to compete in the second round. In the following, we focus on how relative performance beliefs affect the decision to enter the first round of the multistage contest. In particular, we check whether participants with intermediate guessed ranks in the baseline tournament are more sensitive to quotas than those with either high or low relative performance guesses.

Figure 4 shows the distribution of rank guesses for Round 0 (left axis), and for each guessed rank the share of participants who select the tournament option in Round 1 (right axis), by treatment condition and gender. As a reminder, a guessed rank of 1 (4) corresponds to the belief of being in the top (bottom) 25% of participants in a given session in terms of Round 0 scores. Remember that the AA condition could not affect Round 0 outcomes in any way as gender quotas were not even mentioned until the beginning of Round 1. Rank guesses, however, were elicited at the end of the experiment, after participants have already experienced the different treatments, so their self-assessment could potentially be biased even if actual performance was the same. Reassuringly, we find no signs of

21. Although the share of female entrants is highest in the session with the lowest share of women, all our results remain qualitatively unchanged if we exclude observations from that session—calculations are available from the authors on request.

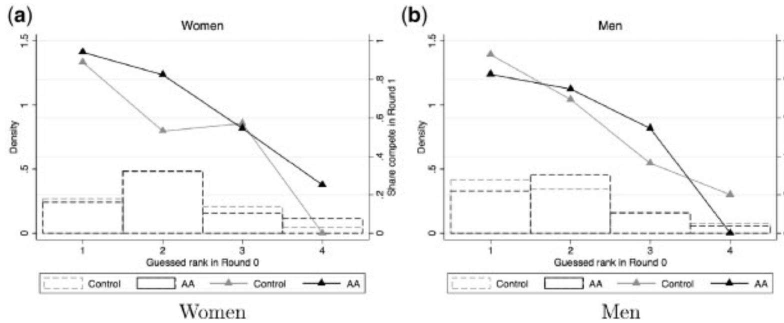


Figure 4. Distribution of Rank Guesses and Their Impact on Tournament Entry.

such bias: there are no significant differences in baseline relative performance guesses between the two treatments.[22] Note that higher relative performance beliefs tend to be associated with greater willingness to compete, lending further credit to our measure of confidence.

Figure 4a suggests that female participants' response to the treatment is indeed heterogeneous with respect to estimated relative performance. Women with a guessed rank of 2 (i.e., those who believe they are among the best 25–50% of participants in the session) are almost 30 percentage points more likely to enter Round 1 tournament in the AA than in the Control condition, while we observe no such treatment effect among women with guessed ranks of 1 or 3. Women who expect to be the worst performers also seem very responsive to the gender quotas; note, however, that very few female participants fall into this category (only three subjects in the Control and eight in the AA treatment). A two-sample t-test with unequal variances confirms that treatment differences in the propensity to compete in Round 1 are highly significant among women with a guessed rank of 2 (p-value$\,=0.011$).[23]

Studying tournament entry rates by treatment and guessed rank among male participants (Figure 4b), we find no support for our hypothesis that quotas disproportionately discourage men with intermediate performance beliefs from competing. In the male subsample, treatment differences in the willingness to compete in Round 1 are relatively small and insignificant irrespective of relative performance beliefs. If anything, it is men with the best and worst guessed ranks who seem to enter Round 1 tournament less in response to AA policies.

22. While male participants in the Control condition were slightly more (less) likely to report a guessed rank of 1 (2) than their peers in the AA treatment, a Kolmogorov–Smirnov test does not reject the equality of the two distributions (p-value$\,=0.948$). In any case, the difference is most likely attributable to true performance differences between the groups stemming from imperfect randomization and unrelated to the specifics of the treatment.

23. The difference in the willingness to compete among women with a guessed rank of 2 in response to the AA condition remains significant after controlling for Round 0 scores. Calculations are available from the authors upon request.

We continue by reviewing potential reasons why quotas leave male participants' decision to compete largely unchanged not only in Round 1 (where they merely affect the option value of future competition) but also in Round 2 (where they directly influence how winners are selected). One explanation is related to confidence: those men who believe they belong to the top performance quartile of tournament entrants (both among their own gender and overall) expect to win the second stage tournament irrespective of the AA policies. Indeed, over 55% of male participants who choose to select the tournament option in Round 2 of the AA treatment believe they belong to the top 25% of entrants (both male and overall) in their session and thus will win the tournament despite the quotas being in place.

Second, whether quotas are perceived as binding depends on participants' perception of gender differences in task performance. For those who expect at least half of the winners to be female even in the absence of AA policies, reserving 50% of winner positions for female entrants does not substantially change their own perceived winning chances.[24] Comparing guessed ranks elicited in different scenarios, we find that the majority (close to 84%) of men expect the same or worse rank when compared with all entrants than when compared with male entrants only, suggesting that they do not consider men to be tougher opponents than women. Finally, men may not be discouraged by the quotas if they believe that the share of female entrants in Round 2 will be very low such that winner positions reserved for women will eventually become available for all entrants to fill (remember that in our experimental instructions we explicitly mention this possibility, see Section 2). However, only 18.6% of male participants in the AA condition guess that the share of female entrants is less than one-fourth (indicating a belief that at least some of the initially reserved winner positions will open up for male competitors), and the share of those with such beliefs is even lower if we restrict our attention to men who chose the tournament option in Round 2 of the AA treatment. We thus conclude that this explanation is unlikely to apply in our setting.

Result 3: The AA treatment encourages women with intermediate guessed ranks to enter the tournament. Relative performance beliefs do not affect male participants' response to the quotas. Men's insensitivity to quotas may be attributed to their high relative performance beliefs, as well as them expecting women to be as good as men in the task.

4.3 Evaluating the AA Policy

In the previous subsections, we have shown that AA policies in the second stage increase tournament entry rates among high ability women in both rounds of the multistage elimination contest without significantly

24. This reasoning is particularly applicable to our setting where comparisons are made on the session level instead of in groups of four to six participants, so the role of noise is much more limited.

decreasing men's willingness to compete, and argued that women's response is moderated by their relative performance beliefs. In the following, we directly examine the impact of quotas on female representation and on efficiency. We also assess whether quotas affect participants' choices in a dictator game.

4.3.1 Gender Composition. Figure 5 presents the gender composition of entrants (who selected the tournament option and had been winners in Round 1) and winners in the second round of the tournament in the Control and AA treatments. There is an overall increase in the number of entrants as a result of the quotas: even though fewer men enter the second stage of the tournament in the AA treatment, the increase in the number of female entrants more than offsets this reduction. In the AA treatment, the pool of entrants is more balanced along gender lines than in the Control treatment. Similarly, the AA treatments lead to an increase in the number of female winners at the cost of male winners. Interestingly, even though the quota only reserves half of the winner positions for women, the share of women among all the winners is more than 50%. When interpreting these results, keep in mind that there were more high performing men than women in the Control, but not in the AA treatment, so these raw comparisons do not speak directly to the fairness of the outcomes.

While Figure 5 provides an overview of female representation—an outcome policy makers might find interesting in and of itself—it is not informative of efficiency as it does not discuss the task scores of entrants or winners. In the following subsection, we discuss different ways of evaluating efficiency in our experiment.

4.3.2 Efficiency. To understand the broader consequences of the AA policy, there are four potential sources of inefficiency to consider: (1) in the absence of quotas, high ability women shy away from competing; (2) in the presence of quotas, high ability men are discouraged from competing; (3) in the presence of quotas, lower performers are promoted to fulfill the quota requirements; and (4) quotas induce lower performance. In the previous section, we have already addressed the first two points by showing that quotas attract high-ability women to tournaments they would otherwise shy away from, and that the increase in female entry is not coupled with a significant decrease in male participants' propensity to compete. We continue by discussing the third and fourth points.

Due to preexisting ability differences between participants in the Control and AA conditions, a simple comparison of winners' scores in the second stage would be misleading. Instead we gauge whether the best performers are announced as winners of Round 2 tournament. In particular, we create a subsample of "top baseline performers" by including from each session the n participants with the highest Round 0 score, where n corresponds to the number of Round 2 winners in that session (remember

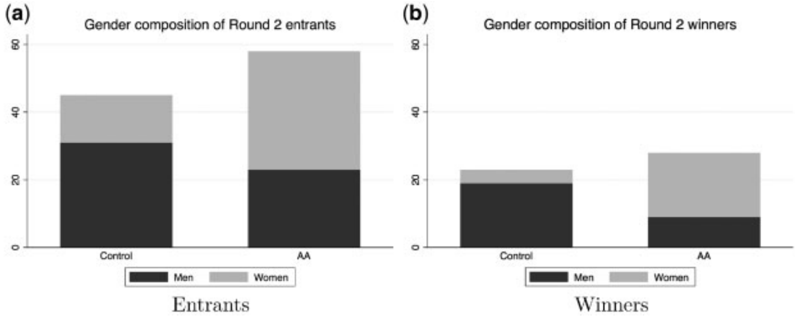


Figure 5. Composition of Tournament Entrants and Winners, Round 2.

that n is endogenously determined by session size and entry choices). In a similar fashion, we also identify the "top final performers," that is, those with the n highest Round 2 scores. We then analyze the probability of becoming a winner in Round 2 among top performers by treatment and gender. The higher this probability, the more efficient a given scheme is in allocating winning positions to the best performers.[25]

Table 4 presents results from this analysis. As Columns (1) and (3) show, AA does not lower efficiency overall: the AA treatment has a negligible and insignificant impact on the likelihood of top performers to become winners in the final round, with estimated coefficients associated with the AA treatment of 0.011 (SE = 0.144) and 0.045 (SE = 0.130), depending on the definition of top performers (in terms of Round 0 and Round 2 scores, respectively). Unsurprisingly, the effect of quotas on top performers' winning chances differs by the gender of the top performer. Columns (2) and (4), displaying results from models that include interaction terms between treatment and gender, show that high-performing women are substantially more likely to be identified as a winner in the AA treatment than in the Control, while top men's chances are somewhat reduced.

To complement the above analysis, we can also count the exact number of cases when the highest performers in the second stage were different than the winners. We are also interested to see whether such instances in the AA treatment were due to participants choosing not to compete, or due to higher performing men being passed over to fulfill the quota requirements ("reverse discrimination"). There were 21 winners in Round 2 in the Control and 28 in the AA treatment. In the Control condition, four participants (three of them women) had higher Round 2 scores than the winners in their sessions, but were not promoted because they did not enter the tournament. In the AA treatment, we identified 10 cases where

25. For a similar overview that includes the full sample, not only the best performers consider Figure A6 in Appendix A plotting the probability of becoming a winner against each participant's within-session relative performance in terms of Round 2 scores, by treatment and gender.

Table 4. Likelihood of Top Performers to Win the Final Stage of the Tournament

	Top baseline performers		Top final performers	
	(1)	(2)	(3)	(4)
AA	0.011	−0.328*	0.045	−0.241
	(0.144)	(0.167)	(0.130)	(0.147)
Female		−0.639***		−0.465**
		(0.211)		(0.197)
AA * Female		0.957***		0.827***
		(0.266)		(0.248)
Score Round 0	0.052**	0.037*	0.035**	0.030**
	(0.024)	(0.022)	(0.016)	(0.014)
Constant	−0.178	0.228	0.195	0.384
	(0.426)	(0.401)	(0.279)	(0.262)
N	51	51	51	51
R^2	0.110	0.306	0.099	0.278

*The table displays OLS estimated coefficients from linear probability models. The dependent variable is a binary indicator for being selected as a winner in the Round 2 tournament. Analysis in Columns (1)–(2) is restricted to the subsample of top performers according to their Round 0 scores, while in Columns (3)–(4) it is restricted to the subsample of top performers according to their Round 2 scores. (See main text for details on determining top performers.) AA denotes the AA treatment, AA Female is the interaction term. Standard errors are in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

a non-winner scored strictly higher than the winners in the session; 8 of these cases concerned a male non-winner with higher performance. There was one single instance when a man who entered the competition was passed over to promote a woman with a lower performance; in the remaining cases the inefficiency stemmed from high ability men forgoing entry.[26] Overall, both in the Control and in the AA condition in more than 90% of the cases participants were "correctly" categorized as non-winners, meaning that they did not perform better in Round 2 than the winners in their session did.

Besides their impact on sorting decisions, quotas could also harm efficiency by lowering participants' effort provision and performance. Intuitively, this could occur if women in the AA treatment anticipated that quotas would help them win with lower scores and thus withheld effort, or if men concluded that quotas decreased their chances of winning to the extent that it was pointless to even try to exert effort.[27] Moreover, gender quotas may impair female performance through a different channel,

26. Given that there was no baseline gender gap in scores in this treatment, these result speak to the efficiency of quotas in case men and women are on average of the same ability, but make different entry choices in "standard" tournaments. Admittedly, it does not tell us how the quota would have performed if applied to our Control sample where men happened to be on average higher performers. We argue that the first situation (no gender gap in performance) is more relevant for the discussion around real-life AA policies.

27. Note that in our setting prizes are not fixed: under both payment schemes, earnings are increasing in the number of tasks correctly solved, possibly limiting the influence of quotas on

by inducing stereotype threat (Bracha et al. 2013). Therefore, in the following we test whether task scores in the first and second round of the multistage tournament differ between the two treatments, by gender. To avoid issues arising from endogenous selection into payment schemes, we analyze scores among *all* participants, not only among tournament entrants or winners. To correct for initial ability differences between participants in the two treatments, we control for Round 0 scores in the regressions.

Table 5 shows that average task performance in both Rounds 1 and 2 of the tournament was unaffected by AA. The estimated coefficients associated with the AA treatment in Columns (1) and (3) are insignificant and small in size: controlling for baseline scores, performance was 0.097 points (SE = 0.288) lower in the first round of the tournament when quotas were in place than in the Control group, while the point estimate is even positive (0.203, SE = 0.307) in the second round. There was no significant gender difference in the performance response, either: as the interaction terms in Columns (2) and (4) show, reaction to the quotas was not substantially different than men's. This finding suggests that stereotype threat did not play a major role in shaping female behavior in this setting.

Result 4: The AA treatment only resulted in one single case of "reverse discrimination" where a higher performing male entrant was passed over to fulfill the the gender quota's requirements. The quotas did seem to discourage a few high performing men from competing who could have won Round 2 tournament had they entered. Quotas did not, however, reduce overall task performance.

4.3.3 Other-Regarding Preferences. We conclude our analysis by considering whether exposure to AA policies affects subsequent other-regarding behavior. In particular, we compare between the two treatments the shares participants assign to the recipient in the dictator game.

Figure 6 reveals a surprising pattern: while women's allocation choices are very similar across treatments, men in the AA condition display much more altruistic behavior. As we have already mentioned in Section 3, the mean amount sent to the receiver is significantly higher in the AA treatment among men, and a Kolmogorov–Smirnov test rejects the equality of the two distributions at a p-value of 0.012.[28] Observing the distribution of the amount sent we find that male participants in the AA treatment are much less likely to keep the full endowment and substantially more likely to choose an equal split. This result is consistent with the explanation that AA treatment made the idea of equity more salient, causing participants to choose "fair" rather than selfish allocations.

performance. For a more systematic discussion on the impact of AA policies on effort provision, please refer to, for example, Schotter and Weigelt (1992).

28. The difference remains substantial and significant even after controlling for task performance, preferences, and beliefs, as well as demographic characteristics. Calculations are available from the authors on request.

Table 5. The Impact of AA on Task Performance

Outcome	Score Round 1		Score Round 2	
	(1)	(2)	(3)	(4)
AA	−0.097	−0.297	0.203	0.100
	(0.288)	(0.410)	(0.307)	(0.438)
Female		−0.696*		−0.426
		(0.415)		(0.444)
AA * Female		0.378		0.193
		(0.575)		(0.615)
Score Round 0	0.883***	0.880***	0.828***	0.826***
	(0.035)	(0.035)	(0.037)	(0.037)
Constant	1.918***	2.378***	2.785***	3.067***
	(0.408)	(0.491)	(0.435)	(0.525)
N	274	274	274	274
R^2	0.716	0.720	0.658	0.659

*The table displays OLS estimated coefficients from linear regression models. The dependent variable is the score in the real effort task in Round 1 (Columns 1–2) and Round 2 (Columns 3–4). AA denotes the AA treatment, AA Female is the interaction term. Standard errors are in parentheses.
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

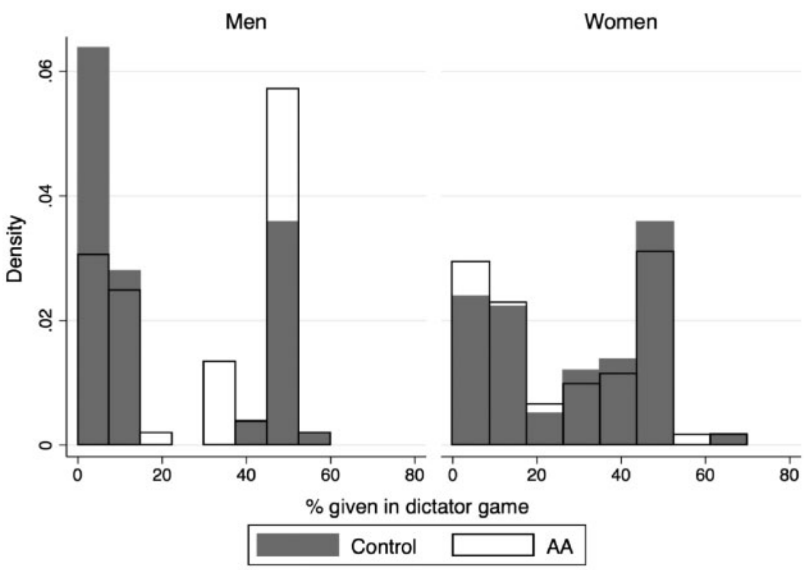


Figure 6. Distribution of Share Allocated to Receiver in Dictator Game, by Treatment and Gender.

5. Discussion

In this section, we discuss the external validity of our findings. Given that we collect our data in a laboratory experiment, we must consider whether the context, the participant pool, and specific experimental design choices limit our ability to generalize our results beyond the particular environment in which they were obtained.

Laboratory experiments grant high levels of control to the researcher over elements of the decision environment, but potentially at the cost of lower external validity (e.g., Harrison and List 2004; Czibor et al. 2019). Studying our research question in the laboratory allows us to eliminate by design several confounds that may affect participants' choices in a real-life promotion contest (e.g., managerial bias, gendered division of housework and childcare, different tasks, etc.). This property of the laboratory reduced the noise, and thus the required sample size, in the experiment. Moreover, it may have been controversial and difficult to implement this experiment if people's real-life career outcomes were at stake. Furthermore, our participants are college students, and as such, not entirely unrepresentative of the population of young professionals who may face gender quotas later in their careers. On the other hand, the context is artificial: choices are made privately (though participants may feel scrutinized, taking part of an overt experiment), there is no social interaction between the contestants, incentives are relatively small, and the task is abstract and repetitive. In the end, the fact that results from related laboratory experiments were successfully replicated in a field setting makes us hopeful that the laboratory context of our study does not seriously compromise the generalizability of our findings.[29]

There are certain elements of our experimental design that deserve further discussion.[30] First, our "multi-stage" contest consists of only two stages. This choice is primarily motivated by practical concerns, such as keeping our experiment short and easy to explain. Two stages are sufficient to demonstrate that the increased option value of future competition due to AA policies can attract women to the tournament even in an early stage without quotas. Our design can thus help alleviate the fear that women's risk aversion and lack of confidence makes them ignore future-stage benefits from AA. However, adding more stages to the tournament, thereby increasing the hurdles women need to pass before reaching the round with the quotas, intuitively reduces the option value of continued competition in the early stages. Future research is needed to determine how many rounds we can add before the beneficial impact of quotas on female entry disappears.

Second, in our setting the two stages of the tournament follow each other directly—therefore, time preferences do not affect participants'

29. Ibanez and Riener (2018) have shown in three real-life hiring experiments that AA policies attract talented female applicants in a one-shot tournament setting, reproducing the laboratory results of BS2012 and NSV2013.

30. We thank Raffaella Sadun, editor of this journal, and two anonymous reviewers, for drawing our attention to these issues.

decisions to compete. In real life, however, later-stages quotas only affect future payoffs, whereas costs and returns from early stages are immediate. Given findings that women tend to be more patient than men on average (Prince and Shawhan 2011; Dittrich and Leipold 2014), we expect that women discount payoffs from future rounds less (and thus respond more strongly to a future quota) than men. As such, we conjecture that later-stage quotas in the future have a stronger impact on women than on men, amplifying their advantages in real-life settings.

Third, the number of winners in the second stage of our tournament is endogenous. Specifically, it is determined as 50% of the second-stage entrants (i.e., those who were winners in the first stage and chose to continue competing in the second stage). This design choice ensures that the probability of winning is constant across sessions with potentially different number of participants. It also has advantages for real-life applications, as it avoids the need to fill up too many pre-determined winner positions when entry is lower than expected, but it also ensures that good candidates do not go unrewarded when entry is higher than anticipated (remember that all the second-stage competitors have cleared a hurdle in the first stage already, so they are a positively selected group). On the other hand, in real-life multistage contests, such as in our motivating example of the competition for corporate board membership, the number of winners is typically fixed. Whether our estimates are too optimistic or too pessimistic compared with the fixed-number scenario depends on participants' beliefs about others' entry decisions, a measure we have not elicited in our experiment. However, we find it reassuring that Maggian et al. (2017) find results similar to ours in a related experiment with fixed number of winners.

Fourth, the way we implemented the second-stage quotas may affect the external validity of our results. In particular, we reserved half of the winner positions for women, whereas in reality quotas typically only require a minimum of 30–40% representation of each gender. We chose the share of reserved seats in keeping with the experiments of BS2012 and NSV2013, to ensure comparability of our results with theirs. Encouragingly, our results suggest that the increase in female entry rates in the AA treatment is not fully attributable to objective changes in the probability of winning.[31] As such, we contend that a lower share of reserved seats would still lead to an increase in female tournament entry.

31. Assume that the best 50% of participants in a session get to enter the second stage, and half of them are women. In this case, only the top 25% of all women in the session are guaranteed to win if they enter the second round in the AA condition. However, in this treatment over 70% of women choose to compete in the first stage, and almost all of them indicate that they would like to continue competing, despite the fact that the majority of women expect a relative rank of 2 or lower (on a scale of 1–4, where 1 is the top rank). The result that quotas generate a response among women above and beyond what we would expect from changes in the objective probability of winning was first documented in NSV2012 in a one-shot tournament context.

Finally, our paper models workplace promotion contests as a forced binary choice between an individual piece rate and a tournament option. This particular design is popular among experimental economists who study gender differences in the willingness to compete [see the seminal paper of Niederle and Vesterlund (2007), and the numerous replications and extensions that followed]. However, results obtained from such designs are probably better applicable to professions that are characterized by an "up-or-out" culture (e.g., academia, law firms, consultancies, and the military) than to workplaces where choices are less explicit and more than the two extreme options are available.

6. Conclusion

This paper reports results from a laboratory experiment that studies the willingness to enter a two-stage elimination contest under two different treatments: one that involves a "standard" tournament, and another where gender quotas are in place in the second round, reserving half of the winning positions for female entrants. Despite the fact that quotas leave the first-stage winning probabilities unchanged, we find that high ability women are more likely to compete already in the first round in response to the AA policy. Women are thus responsive to the increase in the option value of future competition possibilities (Altmann et al. 2012). This finding is consistent with the results of Ifcher and Zarghamee (2015) who show that women can be incentivized to compete but require a premium to do so, and of Petrie and Segal (2014) who find that for large enough rewards, women are willing to compete as much as men. We show that this "premium" or higher prize may come in the form of a greater option value of continued competition, induced by quotas that increase the probability of winning and create a more appealing single-sex competition environment for women in the second round.

The response to quotas is primarily driven by women with intermediate guessed ranks. This finding is intuitive, given that those who expect to be low performers are unlikely to make it to the next stage and thus do not expect to benefit from the AA policy, while those who are very confident of their relative performance expect to win even without quotas. Consequently, real life applications of later stage quotas are most likely to succeed in combination with advice (Brandts et al. 2015) or relative performance feedback (Wozniak et al. 2014), ensuring that women are confident enough to respond to the policy.

Our results support the claim that AA policies applied later along the career path, such as gender quotas for corporate boards, attract talented women to the competition already in earlier stages. This finding is particularly promising in light of recent studies showing that AA measures are not necessarily unpopular: Balafoutas et al. (2016) demonstrate in a laboratory experiment that gender quotas are not necessarily imposed top-down but frequently emerge endogenously in contests. A greater focus on quota-based policies is also important given Baldiga and Coffman (2018)'s result that sponsorship programs, another favored approach among companies interested in bringing about gender diversity, do not actually increase female competitiveness. An alternative, rather radical proposal to ensure that women enter races has been put forth by Goodall and Osterloh (2015) who suggest a random selection procedure from a preselected pool. Quotas are, however, more likely to gain popular acceptance than a random selection mechanism.

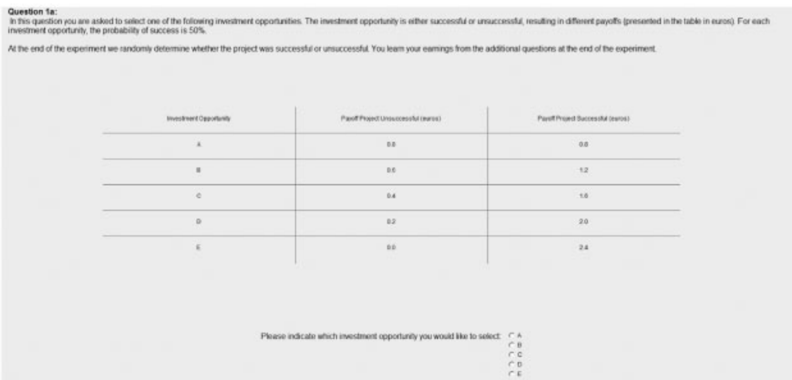
To conclude, we cautiously extrapolate our findings to suggest that the multistage nature of the contest is not the reason why the gender quotas imposed in Norwegian corporate boards left younger women's career-related decisions largely unaffected. However, we acknowledge that real-life decisions to enter promotion contests are influenced by factors that are missing from our experiment. One such factor is the potential for backlash or sabotage against women under an AA treatment (Brown and Chowdhury 2017; Fallucchi and Quercia 2018; Leibbrandt et al. 2018); another is the requirement of costly investment in human capital acquisition early in the game (Cotton et al. 2014; Stark and Hyll 2014). Both of these examples represent interesting avenues for future research to contribute to our knowledge on the determinants of women's labor market outcomes.

Funding

We are grateful to the Research Priority Area Behavioral Economics of the University of Amsterdam for their generous support. E.C. gratefully acknowledges the receipt of the Rubicon grant from the Netherlands Organisation for Scientific Research (NWO) that made her research possible.

Conflict of interest statement. None declared.

Appendix A: Additional Tables and Figures



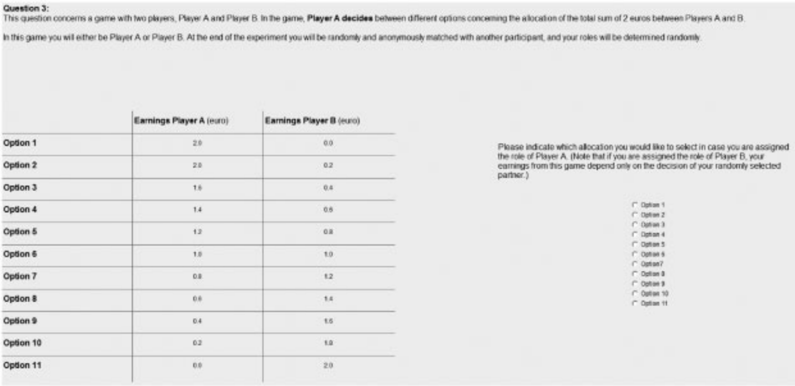
Question 1a:
In this question you are asked to select one of the following investment opportunities. The investment opportunity is either successful or unsuccessful, resulting in different payoffs (presented in the table in euros). For each investment opportunity, the probability of success is 50%.

At the end of the experiment we randomly determine whether the project was successful or unsuccessful. You learn your earnings from the additional questions at the end of the experiment.

Investment Opportunity	Payoff Project Unsuccessful (euros)	Payoff Project Successful (euros)
A	0.8	0.8
B	0.6	1.2
C	0.4	1.6
D	0.2	2.0
E	0.0	2.4

Please indicate which investment opportunity you would like to select: A B C D E

Figure A1. Eliciting Risk Aversion-Incentivized Choice.



Question 3:
This question concerns a game with two players, Player A and Player B. In the game, **Player A decides** between different options concerning the allocation of the total sum of 2 euros between Players A and B.

In this game you will either be Player A or Player B. At the end of the experiment you will be randomly and anonymously matched with another participant, and your roles will be determined randomly.

	Earnings Player A (euro)	**Earnings Player B** (euro)
Option 1	2.0	0.0
Option 2	2.0	0.2
Option 3	1.6	0.4
Option 4	1.4	0.6
Option 5	1.2	0.8
Option 6	1.0	1.0
Option 7	0.8	1.2
Option 8	0.6	1.4
Option 9	0.4	1.6
Option 10	0.2	1.8
Option 11	0.0	2.0

Please indicate which allocation you would like to select in case you are assigned the role of Player A. (Note that if you are assigned the role of Player B, your earnings from this game depend only on the decision of your randomly selected partner.)

Option 1
Option 2
Option 3
Option 4
Option 5
Option 6
Option7
Option 8
Option 9
Option 10
Option 11

Figure A2. Eliciting Other-Regarding Preferences-incentivized Choice.

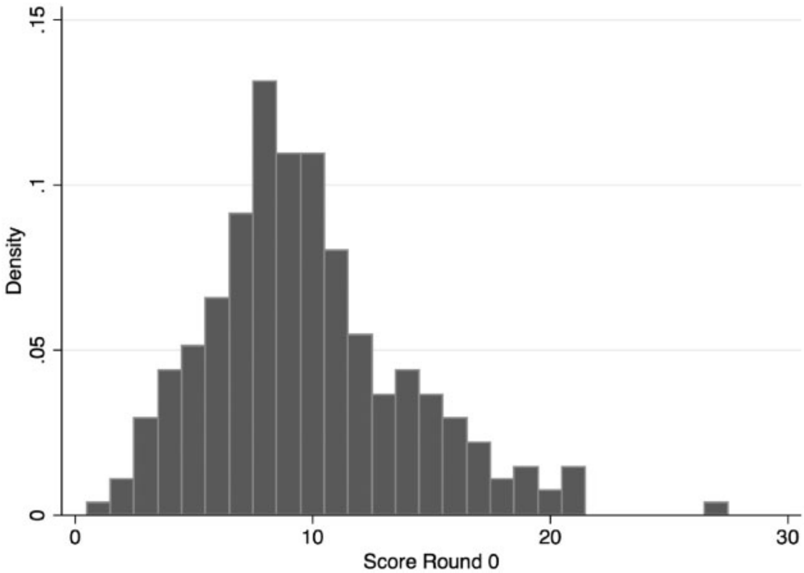


Figure A3. Distribution of Scores in Round 0.

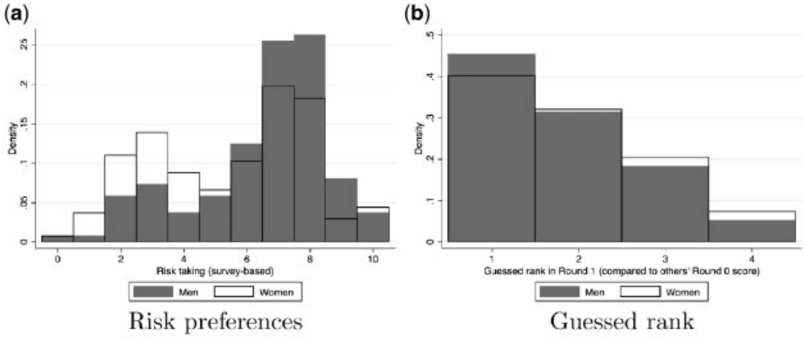


Figure A4. Gender Differences in Preferences and Beliefs—Alternative Measures.

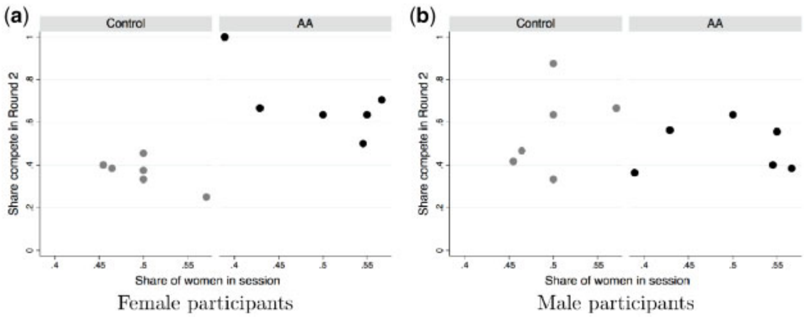


Figure A5. Share Competing in Round 2, by Share of Women in Session.

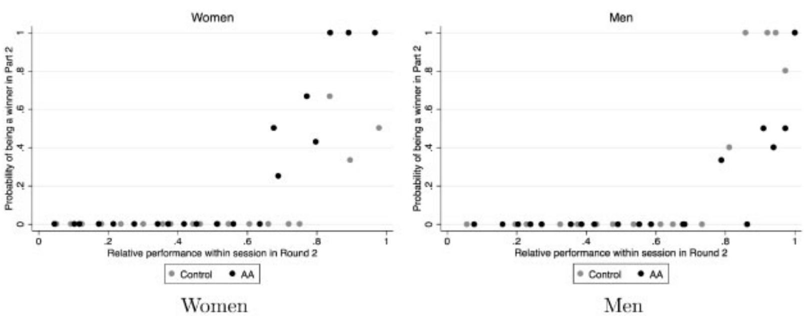


Figure A6. Probability of being a Winner in Round 2, by Within-Session Relative Performance.

Table A1. Overview of the Experimental Sessions

Session number	Treatment	Number of participants	Share of women
1	AA	28	0.43
2	Control	28	0.46
3	Control	22	0.5
4	AA	22	0.5
5	Control	16	0.5
6	AA	20	0.55
7	Control	18	0.5
8	Control	22	0.45
9	AA	22	0.54
10	AA	30	0.57
11	Control	28	0.57
12	AA	18	0.39

Table A2. The Response to AA among Low Ability Participants

	Round 1 (1)	Round 2 (2)
Female	0.125	0.031
	(0.120)	(0.121)
AA	0.132	0.096
	(0.119)	(0.121)
AA * Female	−0.086	0.071
	(0.162)	(0.164)
Score Round 0	0.048**	0.031
	(0.021)	(0.021)
Constant	0.166	0.122
	(0.176)	(0.178)
N	147	147
R^2	0.045	0.032
p(AA+AA * Female=0)	0.681	0.140

The table displays OLS estimated coefficients from linear probability models. The models are estimated on the subsample of participants whose Round 0 task scores are lower than 9. The dependent variable is a binary indicator for selecting the tournament option. AA denotes the affirmative action treatment, AA Female is the interaction term. p(AA+AA * Female=0) presents p-values from F-tests testing whether the sum of the estimated coefficients differs from zero. Standard errors are in parentheses.
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A3. The Impact of Risk Preferences and Confidence

	Full sample		High ability subsample	
	Round 1 (1)	Round 2 (2)	Round 1 (3)	Round 2 (4)
Female	−0.003	−0.101	−0.146	−0.243**
	(0.075)	(0.081)	(0.097)	(0.107)
AA	0.021	−0.031	−0.031	−0.118
	(0.072)	(0.078)	(0.092)	(0.103)
AA * Female	0.122	0.355***	0.281**	0.615***
	(0.102)	(0.110)	(0.133)	(0.148)
Score Round 0	0.000	0.008	0.004	0.012
	(0.008)	(0.008)	(0.012)	(0.014)
Risk taking	0.054***	0.048**	0.065***	0.071***
	(0.019)	(0.020)	(0.024)	(0.027)
Guessed rank Round 0	−0.219***	−0.197***	−0.234***	−0.248***
	(0.037)	(0.039)	(0.062)	(0.069)
Constant	0.896***	0.649***	0.876***	0.631**
	(0.168)	(0.181)	(0.266)	(0.296)
N	274	274	127	127
R^2	0.221	0.228	0.246	0.322
p(AA+AA * Female=0)	0.045	0.000	0.010	0.000

The table displays OLS estimated coefficients from linear probability models. The dependent variable is a binary indicator for selecting the tournament option (column headers indicate the relevant round). Columns (1)–(2) present estimates for the full sample, while the analysis presented in Columns (3)–(4) is restricted to the subsample of high ability participants. AA denotes the affirmative action treatment, AA Female is the interaction term. p(AA+AA * Female=0) presents p-values from F-tests testing whether the sum of the estimated coefficients differs from zero. Standard errors in parentheses.
$^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix B: Experimental Instructions

Welcome to this Experiment

General Information

Thank you for participating in this experiment. In addition to the participation fee of 8 euros, you may earn money by completing tasks and answering questions. The amount you earn depends on the decisions that you and the other participants in the experiment make. Throughout the experiment we ensure strict anonymity: no other participant in the experiment will learn how much money you have earned or how you have behaved in the experiment.

Please do not communicate with other participants during the experiment. If you have a question, please raise your hand. The experimenter will then come to your table to answer your question in private. You are not allowed to use your mobile phone, table, or laptop during the experiment.

Before we begin with the experiment, please answer the following background questions:

What is your gender? 0 Female 0 Male

What is your age?

What is your nationality?
What is your field of study?
What was your high school math grade? (Please use a scale from 1 to 10)
Please click OK when you are ready

Overview of the Experiment

Total earnings from the experiment

The experiment consists of three parts. One of these parts will be randomly chosen at the end of the experiment to be paid. The three parts are followed by additional questions. Your *total earnings* from the experiment are thus the sum of the participation fee of 8 euros, your payments from the randomly selected part, and your earnings from the additional questions.

Experimental Task

In each part, you are required to perform the same task: calculate the correct sum of series of five 2-digit numbers. You have 4 min in each part to work on this task. It is strictly forbidden to use a calculator or your phone to calculate the sum. We refer to the number of problems you solve correctly as your *performance* in the given part.

Example: $23+76+91+33+12=?$

Performance Feedback During the Experiment

At the end of each part you will be informed about the number of problems you solved correctly, but not about your performance when compared with other participants. You will only be informed at the end of the experiment about your relative performance in the part that was randomly selected for payment (when applicable).

Before we begin with Part 1, you get *2 min to practice* the task. *This practice round is not paid. When you are done with reading the instructions, please click OK. The task will start when everybody is ready.*

<Perform Practice round experimental task>

The time is up.

Your total score for the practice round is <Score>

Please remain silent and do not communicate with your neighbours until the next part starts.

Part 1: Tournament

You will be given 4 min to calculate the correct sum of a series of five 2-digit numbers. If Part 1 is randomly selected for payment, then your earnings will depend on the number of problems you solve correctly, when compared with the Part 1 performance of all other participants in this experimental session. *The best-performing 50% of participants will be selected as winners of the tournament.*

In particular, all <N> participants in this session will be ranked based on the number of problems they solved correctly, where 1 corresponds to the best rank and <N> corresponds to the worst rank. Ties are broken at random. You are a winner in Part 1 if your rank is lower than or equal to <N/2>. You receive 1 euro per correctly solved problem if you are a winner and you receive 0 euro otherwise.

When you are done with reading the instructions, please click OK. The task will start when everybody is ready.

<Perform Experimental Task Part 1>

The time is up.

Your total score for Part 1 is <Score>.

Please remain silent and do not communicate with your neighbours until next part starts.

Please read the instructions given to you on paper carefully. When you are done with reading the instructions, please click OK to answer the control questions.

<Distribute Paper Instructions (see Appendix B1 and Appendix B2)>

<Answer Control Questions>

Part 2: Choice between Piece Rate and Tournament Scheme

We now ask you to choose between a *piece rate* and a *tournament* scheme.

Reminder payoffs:

In the tournament scheme, you receive 1 euro per correctly solved problem if you are a winner, and 0 euro otherwise.

In the piece rate scheme, you receive 0.5 euro per correctly solved problem.

Remember, that the tournament option in Part 3 *will* only be available for participants who in Part 2 *choose* the tournament and are winners.

Please choose how you would like to be paid for your performance in Part 2:

❍ Piece Rate Scheme ❍ Tournament Scheme

Please click OK when you are ready. The task will start once everybody is ready.

<Perform Experimental Task Part 2>

The time is up.

Your total score for Part 2 is <Score>.

Please remain silent and do not communicate with your neighbors until next part starts.

Subjects who chose Tournament Scheme in Part 2 see following screen:

Part 3: Choice between Piece Rate and Standard (AA) Tournament Scheme

We now ask you to choose between a *piece rate* and a *standard (AA)* tournament scheme.

Reminder payoffs:

In the standard (AA) tournament scheme, you receive *1 euro* per correctly solved problem if you are a winner, and *0 euro* otherwise.

In the piece rate scheme, you receive *0.5 euro* per correctly solved problem.

Note that you will not be informed about whether you were a winner or not in Part 2 until the end of the experiment. Hence, ALL participants who selected the tournament in Part 2 are now asked to choose a payment scheme for Part 3, but their choice will only be implemented if they were indeed winners in Part 2 and if Part 3 is randomly selected to be paid.

Please choose how you would like to be paid for your performance in Part 3:

❍ Piece Rate Scheme ❍ Standard (AA) Tournament Scheme

Please click OK when you are ready. The task will start once everybody is ready.

Subjects who chose Piece Rate Scheme in Part 2 see following screen:

Part 3: Choice between Piece Rate and Standard (AA) Tournament Scheme

You selected piece rate in Part 2. This means that you are automatically paid according to a piece rate scheme in Part 3.

Reminder payoff:

In the piece rate scheme, you receive *0.5 euro* per correctly solved problem.

Please click OK when you are ready. The task will start once everybody is ready.

<Perform Experimental Task Part 3 >

The time is up.

Your total score for Part 3 is <Score>.

Please remain silent and do not communicate with your neighbors until next part starts.

Additional Questions

Question 1a

In this question you are asked to select one of the following investment opportunities. The investment opportunity is either successful or unsuccessful, resulting in different payoffs (presented in the table in euros). For each investment opportunity, the probability of success is 50%.

At the end of the experiment we randomly determine whether the project was successful or unsuccessful. You learn your earnings from the additional questions at the end of the experiment.

Investment opportunity	Payoff project unsuccessful (euros)	Payoff project successful (euros)
A	0.8	0.8
B	0.6	1.2
C	0.4	1.6
D	0.2	2.0
E	0.0	2.4

Please indicate which investment opportunity you would like to select:
❍ A ❍ B ❍ C ❍ D ❍ E

Question 1b

How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

Please choose a value on the scale below, where the value 0 means "unwilling to take risks" and the value 10 means "fully prepared to take risk."

0 = Unwilling to take risks ❍ ❍ ❍ ❍ ❍ ❍ ❍ ❍ ❍ ❍ ❍ 10 = Fully prepared to take risk

Question 2

This question consists of five subquestions, 2a, 2b, 2c, 2d, and 2e. One of these subquestions will be randomly selected at the end of the experiment to be paid.

Question 2a

We ask you to evaluate your task performance in *Part 1 (Tournament)*, *relative to all other participants in this session*. In particular, we ask you to place your performance in one of the four performance quartiles, where 1 corresponds to the 25% of participants with the highest performance, and 4 to the 25% of participants with the lowest performance.

Your performance in Part 1 was <Score Part 1> correctly solved problems.

If this subquestion is randomly selected at the end of the experiment to be paid, then you earn 1 euro for a correct guess.

Please enter your guess for your rank:

❍ 1 (best 25%) ❍ 2 ❍ 3 ❍ 4 (worst 25%)

Question 2b

We ask you to evaluate your task performance in *Part 2 (Choice between piece rate and tournament)*, *relative to the Part 1 performance of all other participants in this session*. In particular, we ask you to place your performance in one of the four performance quartiles, where 1 corresponds to the 25% of participants with the highest performance, and 4 to the 25% of participants with the lowest performance.

Your performance in Part 2 was <Score Part 2> correctly solved problems.

If this subquestion is randomly selected at the end of the experiment to be paid, then you earn 1 euro for a correct guess.

Please enter your guess for your rank:

❍ 1 (best 25%) ❍ 2 ❍ 3 ❍ 4 (worst 25%)

Question 2c

We ask you to evaluate your task performance in *Part 3*, *relative to the participants who entered the standard (AA) tournament in Part 3 (i.e., participants who were winners in Part 2 tournament and also selected the tournament option in Part 3)*. In particular, we ask you to place your performance in one of the four performance quartiles, where 1 corresponds to the 25% of entrants with the highest performance, and 4 to the 25% of entrants with the lowest performance.

Your performance in Part 3 was <Score Part 3> correctly solved problems.

If this subquestion is randomly selected at the end of the experiment to be paid, then you earn 1 euro for a correct guess.

Please enter your guess for your rank:

❍ 1 (best 25%) ❍ 2 ❍ 3 ❍ 4 (worst 25%)

Question 2d

We ask you to evaluate your task performance in *Part 3*, *relative to the <Gender Participant; female or male> participants who entered the standard (AA) tournament in Part 3 (i.e., <Gender Participant; female or male> participants who were winners in Part 2 tournament and also selected the tournament option in Part 3)*. In particular, we ask you to place your performance in one of the four performance quartiles, where 1 corresponds to the 25% of <Gender Participant; female or male> entrants with the highest performance, and 4 to the 25% of <Gender Participant; female or male> entrants with the lowest performance.

If this subquestion is randomly selected at the end of the experiment to be paid, then you earn 1 euro for a correct guess.

Please enter your guess for your rank:

❍ 1 (best 25%) ❍ 2 ❍ 3 ❍ 4 (worst 25%)

Question 2e

What do you think the *share of women* was among the standard (AA) tournament entrants in *Part 3* (i.e., participants who were winners in Part 2 tournament and selected the tournament option in Part 3)?

If this subquestion is randomly selected at the end of the experiment to be paid, then you earn 1 euro for a correct guess.

Please enter your guess for the share of women:

❍ <25% ❍ 25–50% ❍ 50–75% ❍ >75%

Question 3

This question concerns a game with two players, Player A and Player B. In the game, *Player A* decides between different options concerning the allocation of the total sum of 2 euros between Players A and B.

In this game you will either be Player A or Player B. At the end of the experiment you will be randomly and anonymously matched with another participant, and your roles will be determined randomly.

	Earnings Player A (euro)	Earnings Player B (euro)
Option 1	2.0	0.0
Option 2	2.0	0.2
Option 3	1.6	0.4
Option 4	1.4	0.6
Option 5	1.2	0.8
Option 6	1.0	1.0
Option 7	0.8	1.2
Option 8	0.6	1.4
Option 9	0.4	1.6
Option 10	0.2	1.8
Option 11	0.0	2.0

Please indicate which allocation you would like to select in case you are assigned the role of Player A. (Note that if you are assigned the role of Player B, your earnings from this game depend only on the decision of your randomly selected partner.)

❍ Option 1 ❍ Option 2 ❍ Option 3 ❍ Option 4 ❍ Option 5 ❍ Option 6 ❍ Option7 ❍ Option 8 ❍ Option 9 ❍ Option 10 ❍ Option 11

Appendix B1: Paper Instructions AA Treatment

Overview Part 2 and Part 3

Choosing Your Payment Scheme

Both Part 2 and Part 3 entail a possibility to choose the payment scheme determining your earnings from these parts (in case one of them is randomly selected to be paid). The choice is always between

- a piece rate scheme (where you get paid for each correctly solved addition task) and
- a tournament scheme (the rules differ between the two parts, as explained below).

Note, however, that *the tournament option in Part 3 is only available for participants who in Part 2 chose the tournament and were winners*. If you select the piece rate scheme in Part 2, you are automatically paid a piece rate in Part 3 as well.

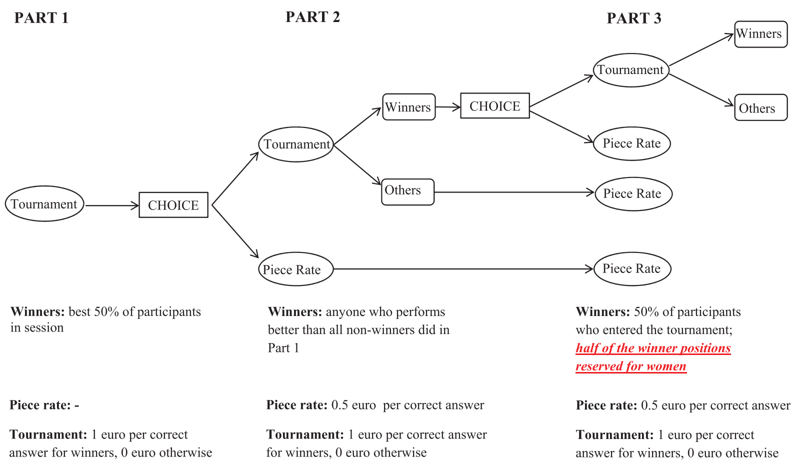
Payoffs under the Two Payment Schemes.

- Under the tournament scheme, you receive 1 euro per correctly solved problem if you are a winner and you receive no payment otherwise.
- Under the piece rate scheme, you receive 0.5 euro per correctly solved problem.

Determining the winners of the tournament. We use different methods to determine the winners of the tournament in Part 2 and in Part 3 (summarized in the figure on the following page):

- Part 2: Competing against others' past performance
  In Part 2, you are a winner if your performance in this part is higher than Part 1 performance of at least 50% of participants in this session. To put it differently, *you are a winner in Part 2 if you solve more problems correctly than all non-winners did in Part 1*. Ties are broken at random.
- Part 3: AA tournament
  In the AA tournament, *50% of those who enter the tournament will win*; half of the winner positions are *reserved for women*, ensuring that *at least 50% of the winners are females* (provided that there are enough female entrants). We use the following procedure to determine who the winners are:
- First, we rank all participants who entered the tournament within their gender group based on their Part 3 performance: females compared with the group of females, males compared with the group of males. (When there are ties in the ranking, they are broken at random.)
- If more female participants enter the tournament than winner positions reserved for women, then the females with the highest rank are placed in the reserved positions. The other female entrants are grouped together with all the male entrants, and a new ranking is determined within this group. The highest ranked participants from this group fill the non-reserved winner positions.
- If at most as many female participants enter the tournament as the number of winner positions reserved for women, then all females who entered are selected as winners. All the remaining winner positions (i.e., reserved but not filled and non-reserved) are filled with the highest ranked male participants.

Note: You will not be informed about whether you were a winner or not in Part 2 until the end of the experiment. Hence, all participants who select the tournament in Part 2 will be asked to choose a payment scheme in Part 3, but their choice will only be implemented if they were indeed winners in Part 2 (and if Part 3 is randomly selected to be paid).



Appendix B2: Paper Instructions Control Treatment

Overview Part 2 and Part 3

Choosing Your Payment Scheme

Both Part 2 and Part 3 entail a possibility to choose the payment scheme determining your earnings from these parts (in case one of them is randomly selected to be paid). The choice is always between

- a piece rate scheme (where you get paid for each correctly solved addition task) and
- a tournament scheme (the rules differ between the two parts, as explained below).

Note, however, that *the tournament option in Part 3 is only available for participants who in Part 2 chose the tournament and were winners*. If you select the piece rate scheme in Part 2, you are automatically paid a piece rate in Part 3 as well.

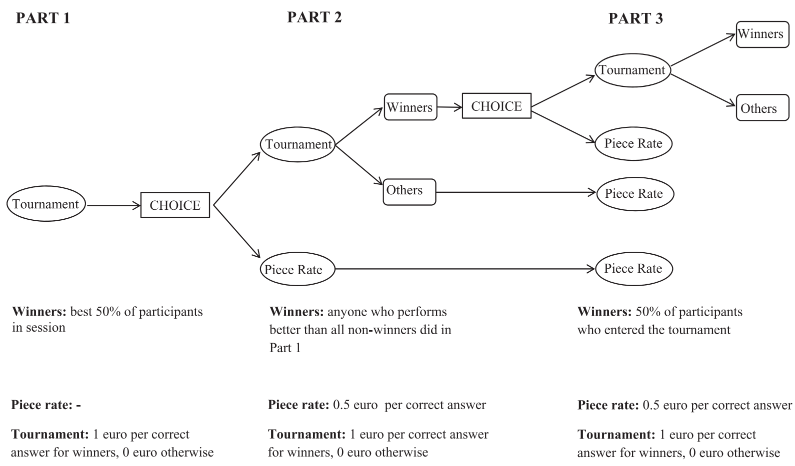
Payoffs under the two payment schemes

- Under the *tournament* scheme, you receive *1 euro* per correctly solved problem if you are a winner and you receive no payment otherwise.
- Under the *piece rate* scheme, you receive *0.5 euro* per correctly solved problem.

Determining the winners of the tournament. We use different methods to determine the winners of the tournament in Part 2 and in Part 3 (summarized in the figure on the following page):

- Part 2: Competing against others' past performance
  In Part 2, you are a winner if your performance in this part is higher than Part 1 performance of at least 50% of participants in this session. To put it differently, *you are a winner in Part 2 if you solve more problems correctly than all non-winners did in Part 1*. Ties are broken at random.
- Part 3: Standard tournament
  In the standard tournament, *50% of those who enter the tournament will win*. We use the following procedure to determine who the winners are. We rank all participants who entered the tournament based on their Part 3 performance. (When there are ties in the ranking, they are broken at random.) The highest ranked 50% of tournament entrants will be placed in the winner positions.

Note: You will not be informed about whether you were a winner or not in Part 2 until the end of the experiment. Hence, all participants who select the tournament in Part 2 will be asked to choose a payment scheme in Part 3, but their choice will only be implemented if they were indeed winners in



Part 2 (and if Part 3 is randomly selected to be paid).

References

Ahern, Kenneth R., and Amy K. Dittmar. 2012. "The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation," 127 *Quarterly Journal of Economics* 137–97.

Altmann, Steffen, Armin Falk, and Matthias Wibral. 2012. "Promotions and Incentives: The Case of Multistage Elimination Tournaments," 30 *Journal of Labor Economics* 149–74.

Azmat, Ghazala, and Barbara Petrongolo. 2014. "Gender and the Labor Market: What Have we Learned from Field and Lab Experiments?," 30 *Labour Economics* 32–40.

Balafoutas, Loukas, and Matthias Sutter. 2012. "Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory," 335 *Science* 579–82.

Balafoutas, Loukas, Brent J. Davis, and Matthias Sutter. 2016. "Affirmative Action or Just Discrimination? A Study on the Endogenous Emergence of Quotas," 127 *Journal of Economic Behavior & Organization* 87–98.

Baldiga, Nancy R., and Katherine B. Coffman. 2018. "Laboratory Evidence on the Effects of Sponsorship on the Competitive Preferences of Men and Women," 64 *Management Science* 888–901.

Baltrunaite, Audinga, Piera Bello, Alessandra Casarico, and Paola Profeta. 2014. "Gender Quotas and the Quality of Politicians," 118 *Journal of Public Economics* 62–74.

Banerjee, Ritwik, Nabanita Datta Gupta, and Marie Claire Villeval. 2016. "The Spillover Effects of Affirmative Action on Competitiveness and Unethical Behavior," 101 *European Economic Review* 567–604.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?," 124 *Quarterly Journal of Economics* 1497–540.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India," 335 *Science* 582–6.

Beaurain, Guillaume, and David Masclet. 2016. "Does Affirmative Action Reduce Gender Discrimination and Enhance Efficiency? New Experimental Evidence," 90 *European Economic Review* 350–62.

Bertrand, Marianne, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney. 2019. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway," 86 *Review of Economic Studies* 191–239.

Booth, Alison L. 2009. "Gender and Competition," 16 *Labour Economics*, 599–606.

Bracha, Anat, Anna Cohen, and Lynn Conell-Price. 2013. *Affirmative Action and Stereotype Threat*. Working Papers 13–14, Working Paper Series, Federal Reserve Bank of Boston.

Brandts, Jordi, Valeska Groenert, and Christina Rott. 2015. "The Impact of Advice on Women's and Men's Selection into Competition," 61 *Management Science* 1018–35.

Brown, Alasdair, and Subhasish M. Chowdhury. 2017. "The Hidden Perils of Affirmative Action: Sabotage in Handicap Contests," 133 *Journal of Economic Behavior and Organization* 273–84.

Buser, Thomas, and Huaiping Yuan. 2016. "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad." Tinbergen Institute Discussion Papers 16-096/I, Tinbergen Institute.

Calsamiglia, Caterina, Jörg Franke, and Pedro Rey-Biel. 2013. "The Incentive Effects of Affirmative Action in a Real-Effort Tournament," 98 *Journal of Public Economics* 15–31.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors," 90 *Review of Economics and Statistics* 414–27.

Charness, Gary, and Uri Gneezy. 2012. "Strong Evidence for Gender Differences in Risk Taking," 83 *Journal of Economic Behavior & Organization* 50–8.

Charness, Gary, Uri Gneezy, and Michael A. Kuhn. 2012. "Experimental Methods: Between-Subject and within-Subject Design," 81 *Journal of Economic Behavior & Organization* 1–8.

Cotton, Christopher, Brent R. Hickman, and Joseph P. Price. 2014. "Affirmative Action and Human Capital Investment: Evidence from a Randomized Field Experiment." Working Paper 20397, National Bureau of Economic Research.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences," 47 *Journal of Economic Literature* 448–74.

Czibor, Eszter, Jimenez-Gomez David, and John A. List. 2019. *The Dozen Things Experimental Economists Should Do (More of). Technical Report, National Bureau of Economic Research*, Cambridge, MA.

Dittrich, Marcus, and Kristina Leipold. 2014. "Gender Differences in Time Preferences," 122 *Economics Letters* 413–5.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jurgen Schüpp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences," 9 *Journal of the European Economic Association* 522–50.

Eckel, Catherine C., and Philip J. Grossman. 2002. "Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk," 23 *Evolution and Human Behavior* 281–95.

———. 2008. "Men, Women and Risk Aversion: Experimental Evidence," in Charles R. Plott and Vernon L. Smith, eds., *Handbook of Experimental Economics Results*, Vol. 1, Part 7, Chap. 113, 1061–73. Handbook: Elsevier.

Fallucchi, Francesco, and Simone Quercia. 2018. "Affirmative Action and Retaliation in Experimental Contests," 156 *Journal of Economic Behavior and Organization* 23–40.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," 10 *Experimental Economics* 171–8.

Fréchette, Guillaume. 2012. "Session-Effects in the Laboratory," 15 *Experimental Economics* 485–98.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians," 90 *American Economic Review* 715–41.

Goodall, Amanda, and Margit Osterloh. 2015. *Women Have to Enter the Leadership Race to Win: Using Random Selection to Increase the Supply of Women into Senior Positions*. IZA Discussion Papers No. 9331.

Grosse, Niels D., Riener Gerhard, and Markus Dertwinkel-Kalt. 2014. "*Explaining Gender Differences in Competitiveness: Testing a Theory on Gender-Task Stereotypes." Working Paper*, University of Mannheim.

Grund, Christian. 2015. "Gender Pay Gaps among Highly Educated professionals—Compensation Components Do Matter," 34 *Labour Economics* 118–26.

Gupta, Nabanita Datta, Anders Poulsen, and Marie Claire Villeval. 2011. "Gender Matching and Competitiveness: Experimental Evidence," 51 *Economic Inquiry* 816–35.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments," 42 *Journal of Economic Literature* 1009–55.

Horrace, William C., and Ronald L. Oaxaca. 2006. "Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model," 90 *Economics Letters* 321–7.

Ibanez, Marcela, and Gerhard Riener. 2018. "Sorting through Affirmative Action: Three Field Experiments in Colombia," 36 *Journal of Labor Economics* 437–78.

Ifcher, John, and Homa Zarghamee. 2015. "Pricing Competition: A New Laboratory Measure of Gender Differences in the Willingness to Compete," 19 *Experimental Economics* 1–21.

Koelle, Felix. 2016. "Affirmative Action and Team Performance." Discussion Papers.

Leibbrandt, Andreas, Liang Choon Wang, and Cordelia Foo. 2018. "Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash against Women," 64 *Management Science*, 3501–16.

List, John A., Azeem M. Shaikh, and Yang Xu. 2016. "Multiple Hypothesis Testing in Experimental Economics." Working Paper 21875, National Bureau of Economic Research.

Maggian, Valeria, and Natalia Montinari. 2017. "The Spillover Effects of Gender Quotas on Dishonesty," 159 *Economics Letters* 33–6.

Maggian, Valeria, Montinari Natalia, and Antonio Nicolò. 2017. *Do Quotas Help Women to Climb the Career Ladder? A Laboratory Experiment*. GATE-Lyon Saint-Etienne, Working Paper, September 19, 2017.

Matsa, David A., and Amalia R. Miller. 2013. "A Female Style in Corporate Leadership? Evidence from Quotas," 5 *American Economic Journal: Applied Economics* 136–69.

Niederle, Muriel, and Vesterlund. Lise 2007. “Do Women Shy Away from Competition? Do Men Compete Too Much?,” 122 *Quarterly Journal of Economics* 1067–101.
Niederle, Muriel, and Lise Vesterlund. 2011. “Gender and Competition,” 3 *Annual Review of Economics* 601–30.
Niederle, Muriel, Carmit Segal, and Lise Vesterlund. 2013. “How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness,” 59 *Management Science* 1–16.
Petrie, Ragan, and Carmit Segal. 2014. “Gender Differences in Competitiveness: The Role of Prizes.” GMU Working Papers in Economics No. 14-47.
Prince, Jeffrey T., and Daniel L. Shawhan. 2011. “Is Time Inconsistency Primarily a Male Problem?,” 18 *Applied Economics Letters* 501–4.
Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. “How Stereotypes Impair Women’s Careers in Science,” 111 *Proceedings of the National Academy of Sciences of the United States of America* 4403–8.
Rosen, Jennifer. 2017. “Gender Quotas for Women in National Politics: A Comparative Analysis across Development Thresholds,” 66 *Social Science Research* 82–101.
Rosen, Sherwin. 1986. “Prizes and Incentives in Elimination Tournaments,” 76 *American Economic Review* 701–15.
Schotter, Andrew, and Keith Weigelt. 1992. “Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results,” 107 *Quarterly Journal of Economics* 511–39.
Stark, Oded, and Walter Hyll. 2014. “Socially Gainful Gender Quotas,” 105 *Journal of Economic Behavior & Organization* 173–7.
Wozniak, David, William T. Harbaugh, and Ulrich Mayr. 2014. “The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices,” 32 *Journal of Labor Economics* 161–98.