CESifo

WORKING Papers

Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash Against Women

Andreas Leibbrandt
Liang Choon Wang
Cordelia Foo

CESifo
Center for Economic Studies & Ifo Institute

Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash Against Women

Abstract

This study experimentally investigates gender quotas in light of peer review. We investigate competitions with and without gender quotas and a peer review process that allows for sabotage. Our findings show that the possibility of peer sabotage renders the gender quota ineffective in encouraging women to enter tournaments and reversing gender pay gaps. Moreover, we provide evidence of a severe backlash against women, as they become targets of sabotage under gender quotas. Interestingly, this is the result of women focusing on sabotaging each other while men sabotage indiscriminately. Our results have implications for the use of quotas to mitigate the under-representation and underperformance of minority groups in environments in which peer sabotage is possible.

Andreas Leibbrandt*
Monash University
Department of Economics
Wellington Road
Australia – 3800 Clayton
andreas.leibbrandt@monash.edu

Liang Choon Wang
Monash University
Department of Economics
Wellington Road
Australia – 3800 Clayton
Liang.C.Wang@monash.edu

Cordelia Foo
Monash University
Department of Economics
Wellington Road
Australia – 3800 Clayton

*corresponding author

This version: 11 September 2015

I.   Introduction

Many seem to share the sentiment that gender quotas are necessary to accomplish what has not happened organically: a greater role for women in business and government. Women are significantly under-represented in leadership positions (Bertrand 2010; Blau et al. 2010). For example, women represent less than 5% of the highest-paid executives in Standard and Poor's datasets, and they occupy less than ⅕ of the board seats of US S&P 500 companies as well as less than ⅕ of the positions in the US legislature. To reduce such gender differences, gender quotas are increasingly applied in many countries. For example, Norwegian companies must adhere to a mandatory 40% quota for female directors. Austrian universities have to re-advertise a position if it has no female applicants and hire female applicants over similarly qualified males (Rasnic, 1995). Recently, Germany has decided that 30% of the board seats of some of the largest companies have to be occupied by women. These measures may have repercussions far beyond their borders and increase the pressure on US companies to implement quotas to increase female representation at higher levels.[1]

There is limited and contradictory evidence on the impact of gender quotas. On the one hand, gender quotas have led to an increase in female representation in leadership positions.[2] In addition, experimental laboratory evidence shows that gender quotas entice women to compete and improve efficiencies (Balafoutas and Sutter, 2012; Niederle et al., 2013). Furthermore, there is evidence that gender quotas help reduce gender-occupation stereotypes (Beaman et al. 2009). Some studies report a positive relationship between women in leadership positions and company performance (McKinsey & Company 2010; Catalyst 2012), although evidence for this has been challenged in more detailed analyses (Adams and Ferreira 2009; Ahern and Dittmar 2012; Matsa and Miller 2013).

On the other hand, there is evidence suggesting that gender quotas are responsible for a *backlash* against women. For example, Ambrose et al. (2002) and Neuman and Baron (1997) suggest that perceived injustice, which could be the result of gender quotas, leads to workplace sabotage and aggression. Beaman et al. (2009) report that gender quotas increase distaste for female leaders. Duffy et al. (2012) find that workers are more likely to undermine their peers when social identification with co-workers is low, which may be the result of differential treatment. Whelan and Wood (2012) show that women who are appointed under gender quotas are regarded as less legitimate, less qualified, and less competent in their roles.

---

[1] http://fortune.com/2015/01/13/catalyst-women-boards-countries-us/

[2] However, the evidence is less clear than one might expect. In Norway, for example, the representation of women in publicly listed companies has increased significantly but at the same time, a large proportion (about 50%) of publicly listed companies has ended public listing to avoid the implementation of gender quotas. See, for example, http://www.sciencedaily.com/releases/2014/02/140224081115.htm.

In this study, we use laboratory experiments to examine whether gender quotas cause a backlash against women and how this affects their pay and willingness to compete. The experimental methodology renders it possible to identify the costs of a backlash clearly and to attribute it to the implementation of a gender quota. We investigate participants performing a real-effort task and their willingness to compete in treatments in which we vary whether there is a gender quota and a peer review. The peer review provides us with the opportunity to capture backlash against women because it provides a channel for sabotage. In our peer review treatments, participants can misreport the performance of others and thereby negatively affect earnings.

Peer review and sabotage are common in many workplaces. Peer review is not just common in academia, but also in governments and firms. For example, the US army relies on Multi-Source Assessment and Feedback, 360-Degrees (MSAF 360) for promotions. Such 360-degree feedback— that is, reviews from members of a worker's immediate work circle—is also widespread in companies and it is estimated that almost 90% of all Fortune 500 companies rely on it (Edwards and Ewen 1996). Survey evidence suggests that 48% of all US workers have either experienced repeated mistreatment or witnessed mistreatment in the workplace, including sabotage, and that women are much more likely to be targeted than are men.[3] Importantly, the availability of peer review and sabotage may affect the willingness to compete and interact with gender quotas. In particular, it may deter women from competing if they expect a higher likelihood of sabotage in the presence of gender quotas.

We find that the impact of the gender quotas crucially depends on the availability of peer review. If there is no peer review, we observe that the gender quota is highly effective in encouraging women to compete, consistent with previous evidence (Balafoutas and Sutter 2012; Niederle et al. 2013). However, if there is peer review, we observe that the gender quota fails to entice women to compete. Moreover, we find that gender quotas make women the prime victims of sabotage. While men's output is more under-reported than women's output if there are no gender quotas, the opposite is true if there are gender quotas. In the latter case, women's output is more than six times more under-reported than men's output. This is because women focus on sabotaging each other in the gender quota tournament while men sabotage men and women equally. Finally, we observe that the availability of peer review implies that gender quotas cannot overcome gender pay gaps.

We view these findings as new evidence on the scope of successfully implementing affirmative action policies. The results offer a novel explanation as to why gender quotas may not be sufficient in encouraging women to compete in workplace environments, and have implications for the use of quotas to secure the representation of minority groups in environments in which peers determine performance. We recommend managers and policymakers should be wary of implementing

---

[3] 2014 WBI U.S. Workplace Bullying Survey February 2014 available at
http://www.workplacebullying.org/wbiresearch/wbi-2014-us-survey/.

gender quotas across the board and avoid peer review when affirmative action policies are present because women become targets of sabotage. More generally, our findings suggest that managers and policymakers should increase their efforts to curb sabotage when implementing gender quotas.

To the best of our knowledge, this is the first study that (*i*) links the research on gender and competitiveness with the research on sabotage and (*ii*) investigates how affirmative action policies are linked to sabotage. Considerable evidence on gender and competitiveness shows that women are less inclined to compete than men in many albeit not all environments (Niederle and Vesterlund 2007; Croson and Gneezy, 2009; Dargnies 2011; Booth and Nolen 2012; Dreber et al. 2012; Brandts et al. 1015; Buser et al. 2015; Flory et al. 2015; Zhang 2015) and there is evidence that sabotage constitutes a serious problem in competitions (Chen 2003; Harbring and Irlenbusch 2005, 2008; Harbring et al. 2007; Carpenter et al. 2010; Schwieren and Weichselbaumer 2010; Vandegift and Yavas 2010; Balafoutas et al. 2012; Gurtler et al. 2013; Charness et al. 2013; Dato and Nieken 2014). However, the latter literature is mainly concerned with the effect of sabotage on effort and does not look at tournament entry in light of sabotage. In turn, the experimental literature on affirmative action has abstracted away from sabotage and focused on environments in which participants cannot directly affect each other (Balafoutas and Sutter 2012; Niederle et al. 2013).

II.   Experiment design

We conducted a laboratory experiment with a total of 100 women and 100 men who took part in one of 10 experimental sessions. In each session, participants were given unique identification numbers and randomly allocated to gender-balanced groups of four; that is, each group consisted of two women and two men. While participants were aware of this process, they did not know who was in their group. We randomly assigned groups to two different treatments: *standard* and *quota*. The experiments were conducted in September 2014 at the Monash University Laboratory for Experimental Economics in Melbourne, Australia.

There were five tasks and participants were informed that they would be compensated for one randomly selected task. This task was selected after participants completed the whole experiment. We did not provide participants with feedback on their own performance during the whole experiment. All participants received a $10 show-up fee. Instructions and question sheets for each task were contained in clearly labeled envelopes on the participants' desks and participants were instructed to open the relevant envelopes only immediately before each task, which they did.

The experiment design builds upon Niederle and Vesterlund (2007). The main novelty in our experiment design is the way in which participants can affect the payment of their group members (Tasks 4 and 5). We create this possibility via a peer review process in which participants can determine objective scores of their group members, which directly translate to payments. They can

accurately over- or under-report scores, and there are financial incentives in place to under-report scores, that is, to sabotage group members to win tournaments. Table 1 summarizes the structure of the experiment for both treatments. The tasks were as follows.

[Insert Table 1 about here]

Task 1—Piece Rate. We asked participants to add up sets of five two-digit numbers for 5 minutes. We gave participants identical question sheets with 40 questions presented in the following way.

| 61 | + | 59 | + | 35 | + | 42 | + | 77 | = | |

Calculators were not allowed but scrap paper was provided. After 5 minutes, we collected the question sheets. The number of correct answers given by participants constituted their Task 1 score. If Task 1 was chosen for payment, participants received $0.50 per question correctly answered. Task 1 was used to measure the ability of participants to perform a simple arithmetic task.

Task 2—Tournament. We asked participants again to add up sets of five two-digit numbers for 5 minutes. In contrast to Task 1, payments in Task 2 depended on relative performance. If this task was chosen for payment, participants received $1 per correct answer if they were a winner of the tournament, and nothing otherwise. The treatments determined how winners were selected. In the standard treatment, tournament winners were the two highest-ranking participants in their group, regardless of gender. In the quota treatment, at least one of the two winners had to be a woman—that is, the treatment implemented a gender quota. Thus, the winners in the treatment were the highest-ranking woman and the highest-ranking group member (of either gender) in the remainder of the group. If there was a tie, there was a random draw.

Task 3—Submit Task 1 Score to Piece Rate or Tournament. After experiencing both the non-competitive compensation scheme in Task 1 and the competitive compensation scheme in Task 2, we asked participants which scheme they wanted to use for their Task 1 score. Thus, as shown in the third row of Table 1, participants in the standard treatment chose between a piece rate and a standard tournament, while participants in the quota treatment chose between a piece rate and a tournament with a gender quota. If Task 3 was chosen for the final payment, participants were paid according to their choice and outcome. Participants received $0.50 per correct answer if the piece rate was chosen. If a tournament was chosen, winners received $1 per correct answer and losers nothing. If a participant selected the tournament, his Task 1 score was compared to the Task 1 scores of *all* his group members, irrespective of the other group members' choices. Thus, we can rule out that gender-dependent beliefs regarding the choices of others affect the willingness to compete. Importantly, this

feature implies that competitions are always gender balanced; that is, two women's and two men's scores are compared to determine the winners.

Task 4—Peer-Review Tournament. Participants were given 5 minutes to add up sets of five two-digit numbers. *Prior* to completing the sums, we told the participants that the scores for Task 4 were not determined by the experimenter but by a peer review during which each participant reviewed the question sheets of his three group members and reported their scores. More precisely, instead of the experimenter counting the number of correct sums, participants determined the number of correct sums. Then, each person's Task 4 score was determined by one randomly drawn peer report. If this task were selected for payment, participants would be paid according to a standard tournament in the standard treatment and a gender quota tournament in the treatment quota. After participants completed the sums, we collected the question sheets and placed them in the front of the room. Participants were then informed that they would complete Task 5 *before* engaging in the peer-review process described.

Task 5—Submit Task 4 Score to Piece Rate or Tournament. We asked participants to select their preferred compensation scheme for their Task 4 score. They were reminded that Task 4 scores would be based on the score in one randomly drawn peer review, which did not have to correspond to the true score (i.e., the actual number of correct answers given). If this task was selected for payment, the participants were paid according to their choices and peer reviewed scores.[4]

Details of the peer review process. After participants completed Task 5, they engaged in the peer-review process to determine their Task 4 scores. We provided each participant with three sheets: *(i)* the three Task 4 question sheets from their group members, *(ii)* a sheet containing the correct answers for Task 4, and *(iii)* a reporting sheet to enter the IDs and scores of their group members. We instructed each participant to fill out the three reporting sheets. The instructions gave examples for reporting sheets, which were filled out correctly and incorrectly (see Appendix).[5] We told participants that we would collect the reporting sheets only during the experiment and that the Task 4 question sheets would remain at their desks for the remainder of the experiment. We collected these sheets only after the participants had left the laboratory and were paid.

---

[4] In contrast to Niederle and Vesterlund's (2007), we asked participants only to select their preferred compensation scheme for a past performance and not for a future performance as well. Doing so and rewarding the task according to peer-reported scores would have introduced complications. First, there would be dependencies between participants' choices and misreporting behavior. For example, a participant who chose to be rewarded according to a piece rate would not have an incentive to misreport others' performance, since his payment would not depend on others' peer-reported scores. Second, effort may be affected also by the choice of compensation scheme. Separating performance and tournament-entry choice into different tasks ensures that participants all perform under the same incentive scheme, and that all participants have the same incentives to misreport others' performance in the peer-review process.

[5] We did not provide a rationale for the peer review and we did not encourage participants to misreport. However, as it was a paper-and-pencil experiment, it seems sensible to assume that participants viewed the peer review process as a means to help the experimenter speed up the calculation process, and thus, the peer review process was embedded in a rather natural environment.

The peer review process entails two important features: the act of sabotage did not incur costs for the perpetuator and sabotage could not be detected. These features account for a broad class of possible acts of sabotage in actual workplace settings (e.g., delaying an action that is supposed to be part of a job, in order to hurt the output of a coworker–competitor). In addition, the features facilitate the interpretation of sabotage and its implications as risk preferences, and gender differences therein, should play no role during the peer review.

In order to convey gender during the peer review process, men's instructions and question sheets inside the envelopes were printed on blue paper, while women's were printed on pink paper. This color procedure was used in both treatments and in all tasks. Participants were informed of the different colored sheets prior to Task 1. This procedure has three advantages. First, by applying it to both treatments, the procedure in itself cannot account for simple treatment differences. Second, by applying it throughout the whole experiment, the procedure in itself cannot account for simple differences across tasks within treatments. Third, this procedure guarantees anonymity while still conveying the gender of each group member. After all five tasks were concluded, participants filled out a short questionnaire, which included a risk questionnaire by Weber et al. (2002) to account for potential gender differences in risk preferences (Croson and Gneezy 2009).

Predictions

In this experiment, participants decide how much effort they make to solve tasks, whether they want to enter tournaments, and the extent to which they misreport during the peer review. The predictions with regard to effort are trivial: participants try to solve as many questions as possible (Tasks 1, 2, and 4).[6] The predictions with regard to willingness to compete, that is, being paid according to a piece rate or tournament (Tasks 3 and 5), depend on the participants' beliefs about their relative performance. In the standard tournament (Task 3), participants should enter the tournament regardless of gender if they believe that their scores are better than the median.[7] However, based on previous evidence, we expect that women are less likely to compete than men in the standard tournament are.

In the quota tournament women have a better chance of winning than do men. Women who believe they are better than the median woman in Task 3 should enter the tournament. Moreover, a woman who believes that the other woman performs better than the other two male group members should enter the tournament, regardless of her own performance. By contrast, men should enter the tournament only if they believe they are better than the median man and the median woman. Thus, we

---

[6] However, in Task 4 if a participant believes that all his group members will misreport and note that he has not solved any questions, there would be no material incentive to solve any question. Such a belief seems not to have played a role, as the number of questions solved is not lower in Task 4 than in the other tasks.

[7] This prediction holds assuming participants are not risk averse. If they believe their performance is equal to the median, they should be indifferent between entering and not entering the tournament assuming risk neutrality.

hypothesize that women are more likely to enter in the quota treatment than in the standard treatment whereas men are less likely to enter in the quota than in the standard treatment. If this is true, the gender gap in the standard tournament significantly shrinks and perhaps even reverses.

The standard economic prediction for misreporting is that participants will maximally under-report, and thus, sabotage their own group members' scores to improve their own chances of winning the tournament in Task 4. This implies that participants should always give their group members a score of zero. However, the experimental evidence on sabotage suggests that such extreme behavior is rare and some individuals sabotage moderately while others even refrain altogether from sabotage (Charness et al. 2013; Flory et al. 2015). In addition, as there is evidence that some individuals are altruistic (Andreoni and Miller 2002) or have preferences for efficiency (Charness and Rabin 2002, Engelmann and Strobel 2004), it could even be that there is over-reporting of scores.

Importantly, there is no apparent reason to assume that sabotage is targeted toward either gender in the standard tournament. However, it seems likely that there is a backlash against women in the quota tournament. Research suggests that women who succeed as a result of affirmative action are likely to be seen as undeserving; moreover, there is evidence that perceived injustice leads to workplace sabotage and aggression (Neuman and Baron 1997; Ambrose et al. 2002; Adams and Ferreira 2009; Ahern and Dittmar 2012; Matsa and Miller 2013). In addition, there is a fear that gender quotas socially divide women who advance regardless of the quota from those who advance because of it. Thus, *our first main hypothesis is that women are more likely to become victims of sabotage than are men in the quota tournament and that they are targeted by men and women*.

Under-reporting reduces the expected payoffs in our setting for participants regardless of their willingness to compete, as participants always compete against all group members. In the limit, this implies that expected payoffs are zero in Task 5 regardless of choice. However, by relaxing the assumption that all participants are perfectly selfish or make no mistakes, *our second main hypothesis is that fewer participants enter tournaments when peer review is introduced* (Task 5 vs. Task 3). The reason is that the potential payoff differences between the piece rate and tournament shrink in the presence of the peer review in Task 5, rendering the tournament less enticing.[8] In addition, *our third main hypothesis is that women's willingness to compete decreases more strongly than men's* does *when peer review is introduced in the quota tournament* because women anticipate that they are targets of sabotage in the quota treatment of Task 5.

---

[8] For example, a participant who is believed to have a score of 16 could obtain either $8 if she chooses a piece rate or $16 if she chooses a competition and wins—a payoff difference between both incentives schemes of *$8*. In the presence of peer review, assuming under-reporting of 50%, a score of 16 translates to earnings of $4 if she chooses a piece rate and $8 if she chooses a competition and wins – ——a payoff difference between both incentives schemes of only *$4*.

III.   Experimental Results

We first report task performance before we study the experimental findings on the willingness to compete with and without peer review. Thereafter, we investigate the victims of sabotage and analyze the characteristics of saboteurs. Table 2 contains descriptive statistics on task performance for each gender. Similarly to Balafoutas and Sutter (2012) and Niederle et al. (2013), we find that men are slightly better in the mathematical summation task. On average, men answered 12.37 questions correctly in Task 1, while women answered 11.1 questions correctly, a difference that is marginally significant at p=0.098 (Mann–Whitney test, two-sided). In Task 2, both genders improve their score, which could be the result of learning or competitive effects. Men have an average score of 13.56 and women of 12.13. The gender difference is marginally insignificant (p=0.101, Mann–Whitney test, two-sided). The average scores for the correct answers for Task 4 are similar to those for Task 2 (13.43 for men, 11.91 for women; p=0.279 Mann–Whitney test, two-sided).

[Insert Table 2 about here]

A.   *Tournament Entry—Willingness to Compete*

Each participant made two tournament entry decisions: one in Task 3, where peer sabotage was not possible, and one in Task 5, where peer sabotage was possible. Figure 2 illustrates the willingness to compete in our four treatments. Each bar contains 50 observations. The upper panel shows the likelihood of tournament entry for men (blue bars) and women (pink bars) in the standard and quota tournaments in the absence of peer review. Consistent with our hypothesis, we find that men are more likely to enter tournaments than are women in the standard tournament (74% vs. 46%; p=0.008 Fisher-exact test, two-sided[9]). Also consistent with our hypothesis, the upper right panel shows that the gender gap reverses when the tournament has a gender quota. In the presence of the gender quota, only 48% of men and 66% of women decide to compete (p=0.106, Fisher-exact test). That is, the gender quota reduces men's willingness to compete (p=0.013) and increases women's willingness to compete (p=0.069).

---

[9] All tests are two-sided unless reported otherwise.
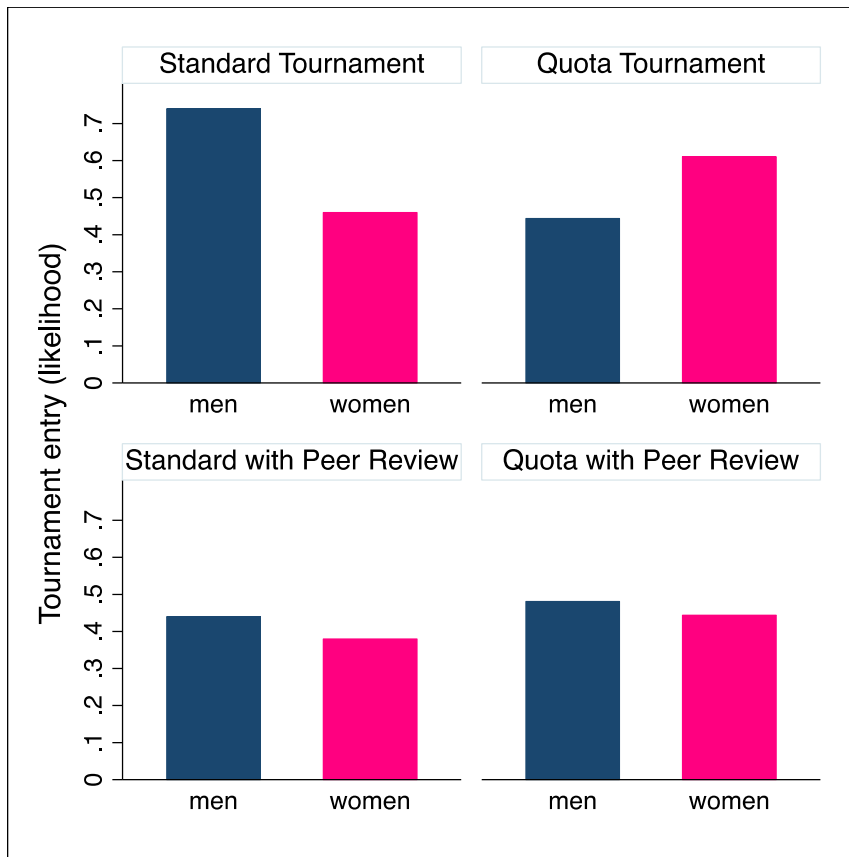
Figure 1: Willingness to compete with and without gender quota and peer review.
The different bars illustrate the likelihood of the willingness to compete in four treatments (standard, quota, standard with peer review, quota with peer review). Each bar contains 50 individual observations. Blue bars are for men, and pink bars are for women. Each participant makes two competition choices, either in the standard or quota tournaments.

The availability of peer review eliminates the power of the gender quota to affect the gender gap in the willingness to compete. The lower panel shows the likelihood of tournament entry when there is peer review. As hypothesized, the willingness to compete is less pronounced overall. The average willingness to compete drops from 59% (no peer review) to 46% (peer review; p=0.012, Fisher-exact test). In particular, in the presence of peer review, men's likelihood to enter standard tournaments drops from 74% to 44% (p=0.002) and women's likelihood drops from 66% to 48% in the quota tournaments (p=0.106). There is no significantly negative impact of the peer review on women's likelihood to enter the standard tournament and men's likelihood to enter the quota tournament. Consequently, the gender quota has no impact on the gender gap in the willingness to compete in the presence of peer review. More precisely, if there is peer review, in the absence of the gender quota, men are 6 percentage points more willing to compete (44% vs. 38%) compared to 4 percentage points more willing to compete in its presence (52% vs. 48%).

Table 3 corroborates and extends the previous findings in six linear probability regression models, which use the willingness to compete in the tournaments with and without peer review as a

dependent variable.[10] We use difference-in-difference regression specification separately by the existence of peer review (i.e., panel A for Task 3 in which no sabotage was possible and panel B for Task 5 in which sabotage was possible). Models 1 and 2 use all participants, Models 3 and 4 use only low-ability participants (103 out of 200 individuals who performed less than the average in Task 1; i.e., scored less than 12), and Models 5 and 6 use only high-ability participants (97 individuals who score 12 or above). Splitting the participants into two groups according to their performance helps to investigate whether gender quotas affect low and high performers' selection in tournaments differently. We are particularly interested in the impact of the gender quota and peer sabotage on the willingness to compete from high-ability women. High-ability women are more likely to win competitions, and successful affirmative action policies need to appeal to those women to increase female representation without sacrificing overall performance. Models 2, 4, and 6, are equivalent to Models 1, 3, and 5 but also control for task performance and risk preferences.

[Insert Table 3 about here]

Model 1 in panel A shows that men are significantly less likely to enter the gender quota tournament than the standard tournament (p<0.01, 26 percentage points) and that women are significantly less likely than men are to enter the standard tournament (p<0.01, 28 percentage points). Importantly, the large marginal effect of 46 percentage points of the gender quota × the female interaction term shows that the gender gap in the standard tournament not only changes significantly but also reverses with the introduction of the gender quota (p<0.01) These findings are robust when controlling for individual task performance and our proxy for risk preferences (Model 2).

The comparison of Models 3 and 5 in panel A indicates that the gender gap in the willingness to enter the standard tournament is larger for the group of high-ability individuals (43.2 percentage points) than in the group of low-ability individuals (13.4 percentage points). Moreover, we observe that the gender quota has a stronger impact on high-ability women (54.6 percentage points) than low-ability women (36.6 percentage points). That is, gender quotas in the absence of the peer review have the desirable feature of reversing the large gender gap among the group of high performers. Models 4 and 6 show the robustness of these findings after controlling for task performance and risk preferences.

For panel B in Table 3, we observe the impact of the gender quota in the presence of peer review. First, in Models 1 and 2, we observe that the gender quota has no significant impact on men's willingness to compete (8 or 6 percentage points, respectively) and that women are not significantly less likely to enter a standard tournaments when there is peer review (−6 or −6.2 percentage points, respectively). Second, we observe that the gender quota does not significantly encourage women more than men to compete when there is peer review (2 or 6.3 percentage points, respectively). Third, for Models 3–6, we observe no clear differences between the low- and high-performing individuals. If at

---

[10] The results are similar if we use logistic regression models. See Appendix II, Table A1.

all, the gender quota seems to encourage low-ability women and discourage high-ability women, but these tendencies are clearly statistically insignificant. Finally, we find that the gender quota × the female interaction term is negative and significantly different between the environment with no peer sabotage and that with peer sabotage. The gender quota × female × the sabotage interaction is −44 to −37 percentage points (p<0.05) for individuals of all ability types, and −59.3 to −53.8 percentage points (p<0.05) for high-ability individuals. Therefore, while gender quotas encourage women to enter tournaments, adding peer review has the opposite effect, and eliminates almost all the gender quota-induced increase in competitiveness, especially among the type of high-ability women that affirmative action seeks to promote.

RESULT 1: *The presence of peer review negates the positive impact of the gender quota on women's willingness to compete. The discouragement effect of the peer review is particularly strong for high-ability women.*

B. Peer Review and its Victims

There is substantial sabotage during the peer review. In Task 4, 43.3% of the peer reviews report scores that are lower than the actual scores. Under-reporting ranges from small to maximal and is roughly equally distributed (12.5% under-report by 1 or 2; 21.3% under-report between 3 and 10, and 9.5% under-report by more than 10) while 10% under-report maximally—that is, report a score of 0. Of the peer reviews, 45.3% are accurate and 11.3% report higher scores than they actually are. On average, scores are under-reported by 12.85%. The tendency to under-report is insignificantly more pronounced in the quota tournament than the standard tournament (14.5% vs. 11.2%, p=0.703, T-test with standard errors clustered at reviewer level). These findings clearly reject the standard prediction that all participants sabotage and maximally under-report.

Who suffers from sabotage? First, overall, we find that women's scores are on average approximately twice as under-reported as men's scores (17.1% vs. 8.6%, p=0.06). Figure 2 illustrates gender differences in being sabotaged depending on the existence of the gender quota. The left panel shows the extent to which scores are under-reported in the standard tournament and the right panel in the quota tournament. Consistent with our hypothesis, we find evidence for backlash against women if and only if there is a gender quota. While women are slightly less sabotaged in the standard tournament (−4.4 percentage points, p=0.39), we observe that they are dramatically more sabotaged in the quota tournament than are men. The mean under-reporting score for women is 25.1% whereas it is only 3.9% for men (p=0.005). Thus, this figure provides the first evidence that women are "punished" for the advantage they gain through the gender quota.
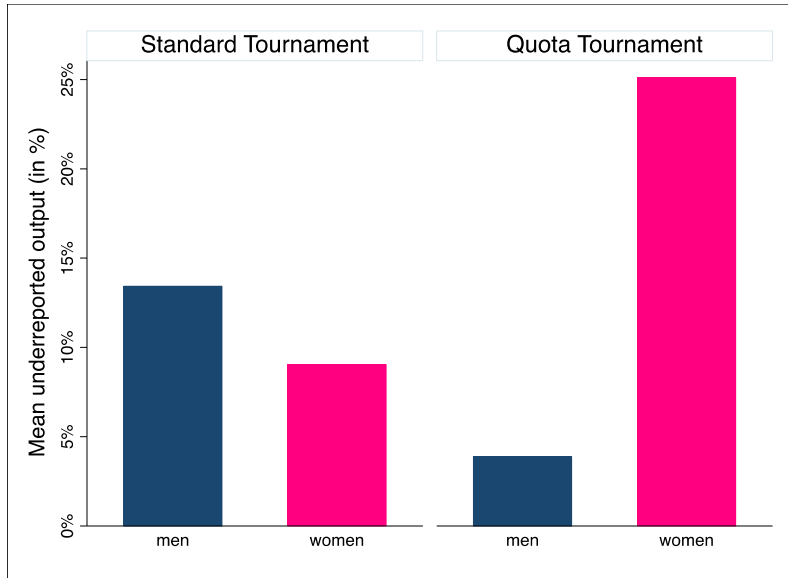
Figure 2: Victims of sabotage with and without gender quota.
The different bars indicate the extent to which output is under-reported in percentages. Blue bars are for men, and pink bars are for women. Each participant receives three peer reviews, that is, each bar contains 150 observations, three per reviewer.

Regression Table 4 corroborates the backlash against women in the presence of a gender quota. We regress misspecification of the score in percentage terms on treatments in six models corresponding to Table 3. There are three observations per participant as each participant is evaluated by her three group members and we cluster standard errors by reviewer. Model 1 in Table 4 presents estimates of specification for the full sample without using controls. The gender quota dummy shows that men's scores are insignificantly less misreported if there is a gender quota. The female dummy shows that women are insignificantly less targets of misreports if there is no gender quota. However, the gender quota × the female interaction term is significantly positive at p<0.01, showing that gender quotas make women targets of sabotage more than men. The coefficient shows that the effect of gender quotas is very large: women's scores are misreported by 25.6 percentage points more compared to men's in the gender quota than in the standard tournament. Model 2 shows that this finding is robust for controlling for task performance and risk preferences (p<0.01). The remaining models show that the previous findings apply to both low and high performers, suggesting that women are targeted uniformly by sabotage but that the negative impact of the gender quota on women's reviewed score is more robust for the group of high performers and significant at p<0.05 in Models 5 and 6.[11]

---

[11] The coefficients of the gender quota × the female interaction term are larger for the low-ability participants (0.284 – 0.299) than for the high-ability participants (0.19 – 0.195). Note, however, that this does not imply that the impact of misreporting is stronger for the low-ability participants. In fact, the impact on high-ability participants is likely to be stronger in absolute terms as they have solved significantly more questions.

[Insert Table 4 about here]

RESULT 2: *The gender quota makes women targets of sabotage. Their scores are more than six times more under-reported than are men's in the presence of a gender quota.*

C.   Peer Review and the Saboteurs

We now examine the origins and targets of sabotage. This analysis allows us to observe if own-gender or cross-gender sabotage is more prevalent, and thus, to understand better how exactly gender quotas exacerbate sabotage. First, there are gender differences in sabotage: 47% of men's peer reviews under-report scores (9.7% over-report, 43.3% are accurate) whereas only 39.7% of women's peer reviews under-report scores (12% over-report, 48.3% are accurate). In particular, men have significantly higher under-reporting scores than women. On average, they under-report by 23.08% whereas women under-report by only 3.82% (p=0.021, t-test, standard errors clustered at reviewer level).[12]

Figure 3 shows the average percentage by which reviewers misreport others' performances during the peer review process depending on the gender of the reviewer (saboteur), the gender of the reviewed (victim), and the existence of the gender quota. This figure provides several new insights into sabotage. First, we observe that male reviewers do not seem to discriminate, as the mean level of under-reporting is similar for male and female victims in both treatments. Second, men increase the level of under-reporting in the gender quota tournament. The mean level of under-reporting is 16.3% in the standard tournament and 29.9% in the gender-quota tournament (p = 0.03, t-test). Third, female reviewers seem to gender discriminate during sabotage. More precisely, they appear to focus on sabotaging men in the standard tournament and women in the gender quota tournament. In the standard tournament, female reviewers under-report male group members' performance by 12.2% and female group members performance by 3.8% (p = 0.045). In the gender-quota tournament, female reviewers over-report male group members' performance by 9.9% but under-report female group members' performance by 14.4%. The gender difference is 24.3% and statistically significant at the 1% level. The over-reporting of men's performance by women is striking and suggests that some women use the peer review mechanism to assist men in overcoming their competitive disadvantage under the gender quota. Together, these patterns explain why women are sabotaged severely in the gender quota tournament and that it is driven by other women and not men.

---

[12] This takes into account all reports, including those with over-reported scores.
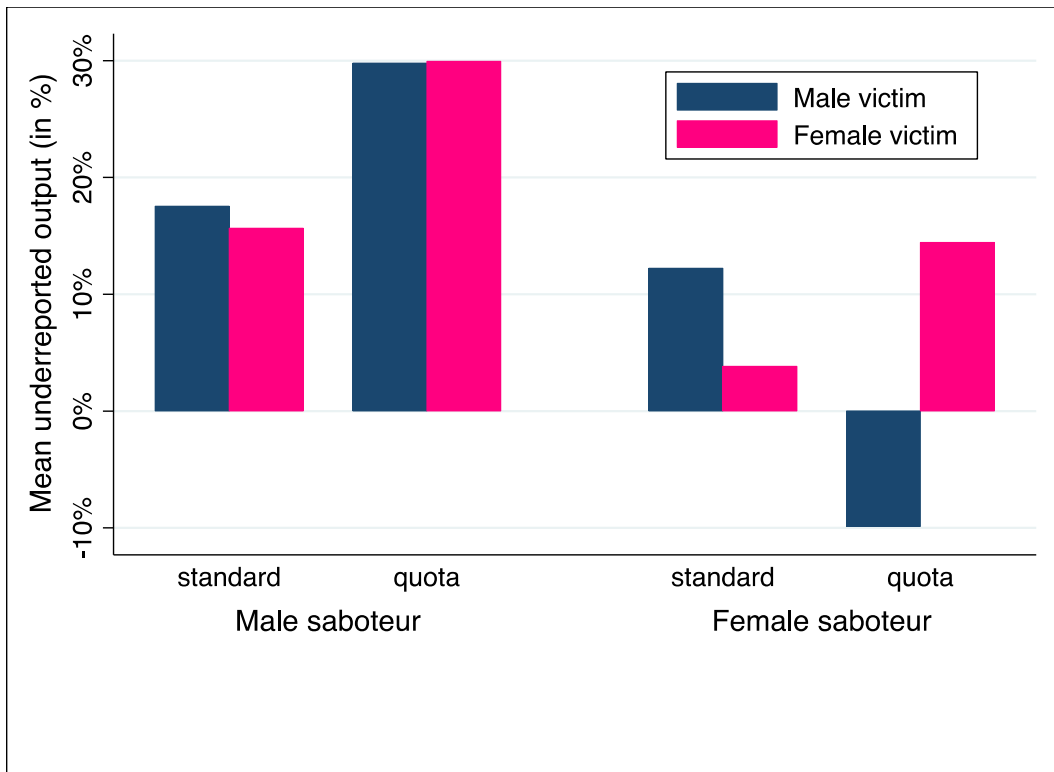
Figure 3: Saboteurs and their victims with and without gender quota.
Blue bars are for male victims, and pink bars are for female victims. Standard = standard tournament, quota = tournament with gender quota. Positive percentages on the y-axis indicate under-reporting of scores (sabotage); negative numbers on the y-axis indicate over-reporting of scores. Each participant receives three peer reviews, that is, each bar contains 75 observations, three per reviewer.

Regression Table 5 provides more in-depth analysis using difference-in-difference regression models to estimate the impact of gender quotas on male and female victims' amount of under-reported performance in the peer-review process by gender of the saboteur. Panel A reports the estimates for male saboteurs and panel B the estimates for female saboteurs. The dependent variable defines the misspecification of the true score in percentage terms. There are three observations per reviewer as each reviewer evaluates three group members. We cluster the standard errors by reviewer. There are six models analog to Table 3.

Models 1 and 2 of panel A show that the extent of under-reporting across treatments is more pronounced if there is a gender quota (12.3 – 12.5%), although the difference is statistically insignificant (p=0.3), and insignificantly less targeted at women if there is no gender quota (2%; p=0.35). Interestingly, Models 3 and 4 show that the gender quota triggers low-ability men to sabotage group members (p=0.06 or p=-.04; 36.7 – 41.3%) but not high-ability men (Models 5 and 6, p=0.46). None of the gender quota × the female victim interaction terms are statistically significant (p>0.68 in all cases) and the coefficients are close to 0, providing clear evidence that men do not sabotage women more than men.

14

Models 1 and 2 of panel B show that female reviewers sabotage women less than men in the standard tournament (p=0.04; 8.4%). At the same time, the gender quota × the female victim interaction terms show that female reviewers sabotage women significantly more than men in the quota tournament (p=0.001; 32.7%). Models 3 and 4 of Panel B show that these patterns are somewhat more pronounced for low-ability women who sabotage men less than women in the quota tournament (p=0.058 or p=0.066, by 48.4 or 44.2%) and men more than women in the standard tournament (p=0.09, 10.9%). However, when low-ability women are in the gender quota tournament, they increase their level of peer sabotage against other group members and do so indiscriminately by gender of their group members (the coefficient of the gender quota × the female victim interaction terms almost undoes the sum of the coefficients of the gender quota and the female victim). Models 5 and 6 show that high-ability women sabotage female group members significantly more when they are in the gender quota tournament, but they do not sabotage their group members much in other treatments, irrespective of the gender of their group members. The bottom row of Table 5 shows that women sabotage other women more than men sabotage women in the gender quota tournament.

[Insert Table 5 about here]

RESULT 3: *Sabotage is more pronounced in the presence of a gender quota. Men sabotage more than women do but do not discriminate between sabotaging men and women. By contrast, women discriminate and focus on sabotaging other women if there is a gender quota.*

D. Gender Gap in Wages

Here, we analyze how quotas affect the gender pay gap. Figure 4 plots women's wages on the vertical axis and men's wages on the horizontal axis. The dotted 45 degrees line illustrates the area where there is no gender gap in wages. We observe nine potential pay outcomes: Task 1, and Tasks 2–5 for each of the standard and quota treatments. The dashed lines connect both treatments to the same task with the arrows showing the direction when moving from the standard to the quota treatment. Thus, the arrows illustrate the impact of the gender quota on each gender's average earnings.

Figure 4 reveals how the gender pay gap is affected by peer reviews and gender quotas. We observe that there is indeed a strong gender pay gap in our tournament (Task 2) and tournament choice (Task 3) condition and that the implementation of a gender quota reverses the gender pay gap. The dashed line connecting the standard and quota tournaments is clearly sloping upwards, showing that the gender quota redistributed earnings between women and men without significantly affecting social efficiency, as defined by joint earnings. However once we allow for selection into competition, we observe a loss in social efficiency. The dashed line connecting standard tournament choice with

15

quota tournament choice is only weakly sloping upwards, showing that little pay increases for women are paired with significant pay losses for men.

Turning to Tasks 4 and 5, which implement the peer review, we observe that gender quotas have little impact on the gender pay gap. There is still a significant gender gap for quota peer review and quota choice peer review, as indicated by the two points on the right side of the dotted no gender-gap line. In addition, we observe that gender quotas in the presence of peer review are at the cost of social efficiency. The line connecting standard with quota peer review is only weakly sloping upwards. If we allow for selection, the corresponding line is even decreasing, showing that both men and women are worse off if a gender quota is implemented in an environment in which peer review is possible. Moreover, the losses for men and women are similar, and thus, the gender pay gap remains similar. Thus, in the presence of peer review, gender quotas can have detrimental outcomes for men and women, and consequently, for social welfare. Appendix II Table A2 supports the previous findings in the two models that estimate the treatment impact on the gender pay gap and average wages.
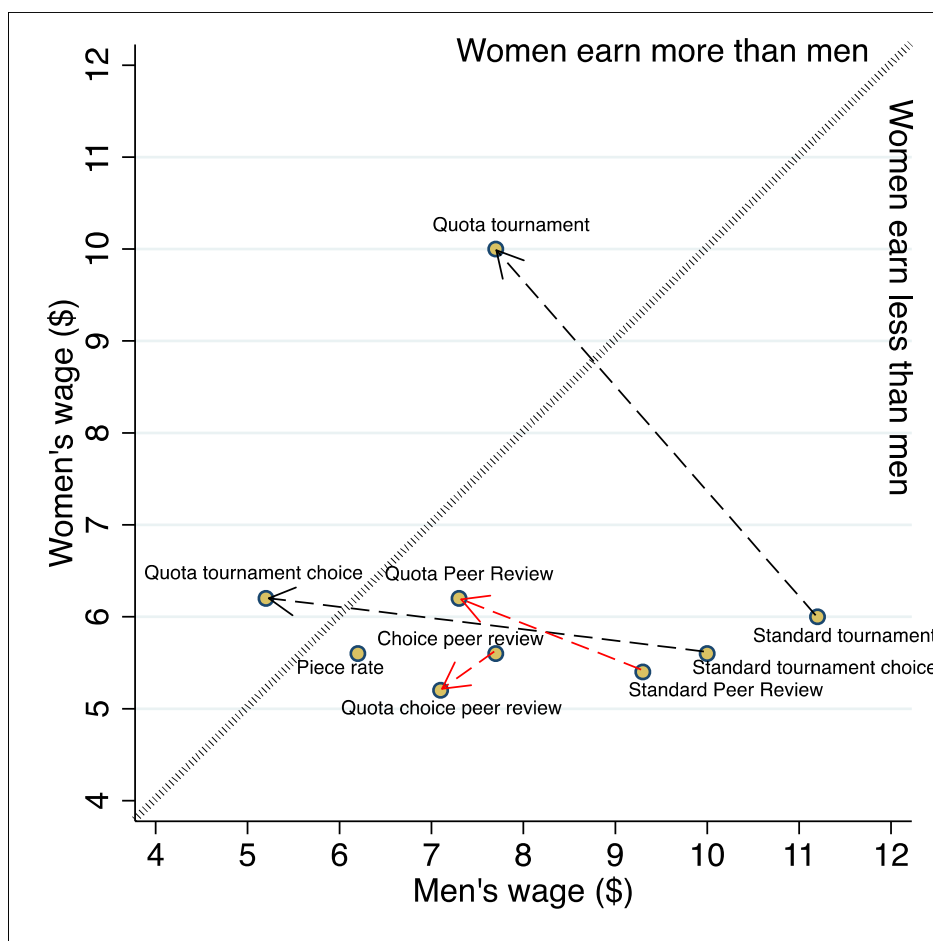
Figure 4: Wages and gender pay gap with and without gender quota and peer review.

The 45° degree line indicates a gender pay gap of zero. Each dot represents one potential pay outcome. The add-on "choice" indicates a task in which participants could select into competition. Piece rate is the baseline Task 1 outcome. The dashed lines connect the corresponding outcomes when moving from the standard to the quota tournaments.

RESULT 4: *In the presence of peer review, gender quotas have little impact on gender pay gaps and reduce social efficiency.*

IV.   Conclusion

Women have been gravely disadvantaged in labor markets. They were not allowed to follow the same professions as men, they have earned substantially less than men have for the same jobs, and they were often bared from career advancement. While many things have changed towards the better in the last decades, women still earn only three-quarters of men's average salaries and are heavily underrepresented at the top of companies and public institutions (Bertrand 2010; Blau 2010). Women still face stereotypes, which render it more difficult for them than for men to bargain, negotiate, and lead (Bowles et al. 2005; Reuben et al. 2012; Bohnet et al. 2014; Reuben et al. 2014). Gender quotas have been proposed and are implemented widely with the aim of closing this gender gap in the hope that they will motivate more women to compete for leadership positions and help remove stereotypes and other resistance.

We use a laboratory experiment to measure the impact of gender quotas on the willingness to compete and sabotage. This methodology comes at a cost—that of investigating the backlash against women in an artificially created decision environment. On the other hand, it allows us to make the backlash visible and estimate the causal impact of gender quotas on the willingness to compete and propensity to sabotage (Falk and Fehr 2003). Our approach is based on previous experimental studies that have been used to show that gender quotas can have desirable impacts when output is measured objectively by a principal. The novelty in our approach is that we allow the economic agents to interact, and thus, to sabotage each other via peer review. Our findings caution against the use of gender quotas in environments in which peer review determines pay. Our study shows that gender quotas do not encourage women to compete when there is peer review. Moreover, we find clear evidence of a backlash as women become the targets of sabotage; they are sabotaged by men and particularly by women. The evidence for the latter is consistent with workplace surveys reporting that women are more than twice as likely to bully other women than men.[13]

---

[13] 2014 WBI US Workplace Bullying Survey, February 2014, available at
http://www.workplacebullying.org/wbiresearch/wbi-2014-us-survey/.

There are several implications to the finding that the possibility of sabotage renders gender quotas ineffective in enticing women to compete and makes them targets of sabotage. Policymakers and managers should be careful in assessing the extent to which opportunities for sabotage exist in their specific contexts and how this may harm women when there are gender quotas in place or planned. For example, the use of "multi-source feedback," in which negative evaluations can have serious financial repercussions for the victim, should be reconsidered. Teamwork may create opportunities for workers to withhold relevant information, which may affect colleagues' output. In addition, access to colleagues' work may provide temptation to destroy such work. Therefore, the extent to which workers are required to engage directly with each other contributes to opportunities to sabotage. Lastly, perhaps the most ubiquitous form of peer sabotage occurs when saboteurs do not affect actual output, but merely attempt to alter perceptions in order to enhance their relative positions. Workers may spread gossip with regard to the quality of colleagues' output, ultimately influencing promotion and remuneration outcomes for their peers. While it may not be possible to eliminate these and other channels of sabotage, it is important to take into account that women may suffer more than men because of the existence of gender quotas.

This study is the first attempt to evaluate the economic costs of a backlash against women caused by affirmative action policies. Our study shows that these costs can be considerable. However, more research is certainly warranted. For example, we study only the impact of a mandatory gender quota but there are several alternatives of affirmative action, such as non-binding targets. In addition, it is important to understand the dynamics of a backlash against women. While there is some evidence that the backlash is not sustained in repeated interactions (Beaman et al. 2009), a vicious circle could exist in which, for example, women respond to other women's sabotage with further sabotage. Lastly, it is important to quantify the role of sabotage in natural environments to formulate concrete policy recommendations.

References

Adams, R. B., and D. Ferreira (2009) Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics* 94: 291–309.

Ahern, K. R., and A. Dittmar (2012) The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics* 127:1 : 137–197.

Ambrose, M. L., Seabright, M. A., & Schminke, M. (2002). Sabotage in the workplace: The role of organizational injustice. *Organizational Behavior and Human Decision Processes*, *89*(1), 947-965.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, 737-753.

Balafoutas L, Lindner F, Sutter M (2012) Sabotage in tournaments: Evidence from a natural experiment. *Kyklos* 65(4):425–441.

Balafoutas L, Sutter M (2012) Affirmative action policies promote women and do not harm efficiency in the laboratory. *Sci.* 335*:*579–582.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: female leadership and gender bias. *Quarterly Journal of Economics*, *124*(4), 1497-1540.

Bertrand M, Hallock KF (2001) The gender gap in top corporate jobs. *Ind. Labour Relat. Rev.* 55(1):3–21.

Bertrand M (2010). New Perspectives on Gender. Card OD, Ashenfelter, O, eds. *Handbook of Labor Economics*, Vol. 4b (Elsevier Science, Amsterdam), 1543–1590.

Blau, F., Feber, M, Winkler, A. 2010. *The Economics of Women, Men and Work* (Prentice Hall, Englewood Cliffs).

Blau F, Kahn LM (2000) Gender differences in pay. *J. Econ. Perspect.* 14(4):75–99.

Bohnet, I., Alexandra van Geen and Max Bazerman. When Performance Trumps Gender Bias. Joint Versus Separate Evaluation. Forthcoming:*Management Science,* 2014.

Booth A, Nolen P (2012) Choosing to compete: How different are girls and boys? *J. Econ. Behav. Organ.* 81:542–555. doi: 10.1016/j.jebo.2011.07.018

Bowles HR, Babcock L, McGinn KL (2005) Constraints and triggers: Situational mechanics of gender in negotiations. *J. Personal. Soc. Psychol.* 89(6):951–965.

Brandts, J., Groenert, V., & Rott, C. (2014). The impact of advice on women's and men's selection into competition. *Management Science*, *61*(5), 1018-1035.

Buser, T., Dreber, A., & Mollerstrom, J. (2015). *Stress Reactions cannot explain the Gender Gap in Willingness to compete* (No. 15-059/I). Tinbergen Institute Discussion Paper.

Buser, T. The impact of losing in a competition on the willingness to seek further challenges *Management Science* forthcoming.

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

Carpenter J, Matthews PH, Schirm J (2010) Tournaments and office politics: Evidence from a real effort experiment. *Am. Econ. Rev.* 100(1):504–517. doi: 10.12507/aer.100.1.504

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly journal of Economics*, 817-869.

Charness, G., Masclet, D., & Villeval, M. C. (2013). The dark side of competition for status. *Management Science*, *60*(1), 38-55.

Catalyst. The Bottom Line: Connecting Corporate Performance and Gender Diversity. New York: Catalyst, 2012. Online at: http://catalyst.org

Chen K-P (2003) Sabotage in promotion tournaments. *J. Law Econ. Organ.* 19(1):119–140.

Croson R, Gneezy U (2009) Gender differences in preferences. *J. Econ. Lit.* 47(2):448–474.

Dargnies M-P (2011) Men too shy away from competition: The case of team competition. *Manag. Sci.* 15(11):1982–2000.

Dato S, Nieken P (2014) Gender differences in competition and sabotage. *J. Econ. Behav. Organ.* 100(2014):68–80.

Duffy MK, Scott KL, Shaw JD, Tepper BJ, Aquino K (2012) A social context model of envy and social undermining. *Acad. Manag. J.* 55(3):643–666.

Edwards, Mark R., & Ewen, Ann J. (1996). *360° Feedback: The powerful new model for Employee Assessment & performance improvement*. New York: AMACOM American Management Association.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American economic review*, 857-869.

Falk A, Fehr E (2003) Why labour market experiments? *Labour Econ.* 10(2003):399–406.

Flory, J. A., Leibbrandt, A., & List, J. A. (2014). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job-Entry Decisions. *The Review of Economic Studies*, rdu030.

Gurtler, O., Munster, J., & Nieken, P. (2013). Information policy in tournaments with sabotage. Scandinavian Journal of Economics, 115, 932-966.

Harbring, C., & Irlenbusch, B. (2005). Incentives in tournaments with endogenous prize selection. *Journal of Institutional and Theoretical Economics (JITE)* 636-663.

Harbring, C., Irlenbusch, B., Kräkel, M., & Selten, R. (2007). Sabotage in corporate contests–an experimental analysis. *International Journal of the Economics of Business*, *14*(3), 367-392.

Harbring, C., & Irlenbusch, B. (2008). How many winners are good to have?: On tournaments with sabotage. *Journal of Economic Behavior & Organization*,*65*(3), 682-702.

Matsa, D., and A. Miller. "A female style in corporate leadership? Evidence from quotas." American Economic Journal: Applied Economics 5:3 (2013): 136–169.

McKinsey & Company (2010) Women Matter—Women at the Top of Corporations: Making it Happen. New York: McKinsey & Company.

Neuman, J. H., & Baron, R. A. (1997). Aggression in the workplace. *Antisocial behavior in organizations*, *37*, 67.

Niederle M, Segal C, Vesterlund L (2013) How costly is diversity? Affirmative action in light of gender differences in competition. *Manag. Sci.* 59(1):1-16. doi: 10.1287/mnsc.1120.1602

Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Q. J. Econ.* 122(3):1067–1101.

Rasnic CD (1995) Austria's affirmative action for women workers versus protective legislation for the "weaker sex": Incongruous concepts? *Labour Law J.* 46(12):749–762.

Reuben, E., Rey-Biel, P., Sapienza, P., & Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, *83*(1), 111-117.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*,*111*(12), 4403-4408.

Schwieren C, Weichselbaumer D (2010) Does competition enhance performance or cheating? A laboratory experiment. *J. Econ. Psychol.* 31(2010):241–253.

Vandegrift, D., & Yavas, A. (2010). An experimental test of sabotage in tournaments. Journal of Institutional and Theoretical Economics, 166, 259-285.

Weber EU, Blais A-R, Betz NE (2002) A domain-specific risk attitude scale: Measuring risk perceptions and risk behaviours. *J. Behav. Decis. Mak.* 15(4):263–290.

Whelan J, Wood R (2012) *Targets and quotas for women in leadership: A global review of policy, practice and psychological research* (Centre for Ethical Leadership, Melbourne).

Zhang, Y J (2015) Culture, Institutions, and the Gender Gap in Competitive Inclination: Evidence from the Communist Experiment in China. *Working paper.*

Table 1: Structure of Experiment

	Standard Treatment	Quota Treatment
Task 1: Piece Rate	Piece Rate	
Task 2: Tournament	Standard Tournament	Quota Tournament
Task 3: Choice 1	Piece rate or standard tournament for Task 1 score	Piece rate or quota tournament for Task 1 score
Task 4: Peer Review Tournament	Standard peer review tournament	Quota peer review tournament
Task 5: Choice 2	Piece rate or standard tournament for Task 4 (peer-reviewed) score	Piece rate or quota tournament for Task 4 (peer-reviewed) score
Peer Review	Peer Review	
Questionnaire	Questionnaire	

Table 2: Descriptive Statistics on Task Performance

	---------- Male ----------				---------- Female ----------				Difference	
Variable	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Mean	p-value
Task 1 score	12.37	5.26	0	38	11.1	4.4	0	23	1.27	0.098
Task 2 score	13.56	5.57	1	34	12.1	4.1	0	24	1.43	0.101
Task 4 answers	13.43	6.66	1	40	11.9	4.3	0	24	1.52	0.279

Notes: The sample size is 100 for each gender. The Task 1 and Task 2 scores are the number of correct answers given in Tasks 1 and 2, respectively. Task 4 answers are the number of correct answers given in Task 4. The *P*-values are based on two-sided Mann–Whitney test statistics.

Table 3: Willingness to compete in tournaments

	(1)	(2)	(3)	(4)	(5)	(6)
	------- All -------		--- Low Ability ---		--- High Ability ---	
A. Non-Sabotage Round						
Gender Quota	-0.260	-0.235	-0.243	-0.220	-0.235	-0.214
	(0.095)***	(0.096)**	(0.145)*	(0.148)	(0.129)*	(0.130)
Female	-0.280	-0.245	-0.134	-0.123	-0.432	-0.405
	(0.095)***	(0.098)**	(0.142)	(0.144)	(0.135)***	(0.140)***
Gender Quota × Female	0.460	0.437	0.366	0.373	0.546	0.530
	(0.137)***	(0.139)***	(0.198)*	(0.198)*	(0.196)***	(0.199)***
Task 1 Score		0.013		0.028		0.018
		(0.006)**		(0.019)		(0.010)*
Risk Score		0.001		0.002		0.000
		(0.001)		(0.002)		(0.002)
Constant	0.740	0.392	0.650	0.126	0.800	0.473
	(0.063)***	(0.207)*	(0.109)***	(0.305)	(0.075)***	(0.345)
B. Sabotage Round						
Gender Quota	0.080	0.060	0.131	0.065	0.065	0.044
	(0.101)	(0.102)	(0.146)	(0.153)	(0.141)	(0.144)
Female	-0.060	-0.062	0.005	-0.033	-0.079	-0.061
	(0.099)	(0.101)	(0.140)	(0.140)	(0.149)	(0.150)
Gender Quota × Female	0.020	0.063	0.034	0.093	-0.046	-0.009
	(0.142)	(0.140)	(0.199)	(0.198)	(0.209)	(0.204)
Task 1 Score		-0.002		-0.044		-0.006
		(0.008)		(0.024)*		(0.013)
Risk Score		0.004		0.003		0.005
		(0.002)***		(0.002)		(0.002)**
Constant	0.440	-0.063	0.350	0.369	0.500	-0.031
	(0.071)***	(0.233)	(0.109)***	(0.365)	(0.093)***	(0.378)
Difference across Rounds in:						
Gender Quota × Female	-0.440	-0.374	-0.333	-0.280	-0.593	-0.538
	(0.183)**	(0.180)**	(0.249)	(0.249)	(0.272)**	(0.261)**
Observations	200	200	103	103	97	97

Notes: Estimates are based on linear probability models. The results are similar when using a logit model (Table A1, Appendix II). Low-ability individuals are those performing below the average score (i.e., 12) in Task 1. Robust standard errors clustered by participant are reported in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

Table 4: Percentage of under-reported scores of victims

	(1)	(2)	(3)	(4)	(5)	(6)
	------- All -------		--- Low Ability ---		--- High Ability ---	
Gender Quota	-0.096	-0.073	-0.181	-0.144	0.016	0.011
	(0.106)	(0.101)	(0.194)	(0.193)	(0.066)	(0.067)
Female	-0.044	-0.017	-0.016	0.003	-0.067	-0.070
	(0.051)	(0.054)	(0.126)	(0.128)	(0.080)	(0.084)
Gender Quota × Female	0.256	0.230	0.299	0.284	0.190	0.195
	(0.089)***	(0.084)***	(0.178)*	(0.183)	(0.095)**	(0.095)**
Task 1 Score		0.010		0.033		-0.003
		(0.006)		(0.017)*		(0.005)
Risk Score		0.000		0.001		0.000
		(0.002)		(0.003)		(0.001)
Constant	0.134	-0.036	0.111	-0.239	0.150	0.161
	(0.056)**	(0.247)	(0.110)	(0.427)	(0.049)***	(0.174)
Observations	600	600	309	309	291	291

Notes: Estimates are based on linear probability models. Low-ability individuals are those performing below the average score (i.e., 12) in Task 1. Robust standard errors clustered by reviewer are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 5: Percentage of under-reported scores of victims by gender of saboteur

	(1)	(2)	(3)	(4)	(5)	(6)
	------- All -------		--- Low Ability ---		--- High Ability ---	
A. Male Saboteurs						
Gender Quota	0.123	0.125	0.367	0.413	-0.112	-0.111
	(0.112)	(0.121)	(0.187)*	(0.197)**	(0.152)	(0.150)
Female Victim	-0.019	-0.019	-0.036	-0.036	-0.007	-0.007
	(0.020)	(0.020)	(0.033)	(0.033)	(0.026)	(0.026)
Gender Quota x Female Victim	0.020	0.020	0.002	0.002	0.051	0.051
	(0.065)	(0.065)	(0.057)	(0.057)	(0.126)	(0.127)
Task 1 Score of Saboteur		-0.000		0.039		-0.003
		(0.010)		(0.032)		(0.013)
Risk Score of Saboteur		-0.001		0.000		-0.002
		(0.002)		(0.003)		(0.003)
Constant	0.175	0.298	0.074	-0.297	0.243	0.528
	(0.079)**	(0.276)	(0.167)	(0.432)	(0.069)***	(0.341)
Observations	300	300	141	141	159	159
B. Female Saboteurs						
Gender Quota	-0.221	-0.216	-0.484	-0.442	0.108	0.108
	(0.147)	(0.147)	(0.250)*	(0.236)*	(0.179)	(0.176)
Female Victim	-0.084	-0.084	-0.109	-0.109	-0.044	-0.044
	(0.041)**	(0.041)**	(0.063)*	(0.063)*	(0.032)	(0.032)
Gender Quota x Female Victim	0.327	0.327	0.467	0.467	0.172	0.172
	(0.098)***	(0.098)***	(0.176)**	(0.177)**	(0.069)**	(0.070)**
Task 1 Score of Saboteur		0.010		0.039		0.015
		(0.016)		(0.037)		(0.022)
Risk Score of Saboteur		0.002		0.003		-0.001
		(0.002)		(0.003)		(0.004)
Constant	0.122	-0.204	0.227	-0.500	-0.048	-0.202
	(0.074)	(0.416)	(0.056)***	(0.702)	(0.167)	(0.475)
Observations	300	300	168	168	132	132
Saboteur Gender Difference in:						
Gender Quota × Female Victim	0.307	0.307	0.465	0.465	0.121	0.121
	(0.117)***	(0.117)***	(0.185)**	(0.186)**	(0.144)	(0.145)

Notes: Estimates are based on linear probability models. Low-ability individuals are those performing below the average score (i.e., 12) in Task 1. Robust standard errors clustered by saboteur (reviewer) are reported in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

Appendix 1: Instructions for "No AA" Treatment (Control)

Page 1

ID: «AddressBlock»

Task 1: Piece Rate

In this task, you will be asked to calculate the sum of five randomly chosen two-digit numbers.

An example is

| 61 | + | 59 | + | 35 | + | 42 | + | 77 | = | ? |

NOTE: You cannot use a calculator to determine this sum, however you may use the scrap paper provided at your desk. Use the pen at your desk to complete the task.

When we tell you to start, you have five minutes to complete as many questions as you can. You may not be able to complete all the questions. At the end of the five minutes, we will collect your question sheets and mark them to determine your Task 1 score, which is the number of questions you answer correctly.

Please continue reading the instructions on this page on your own to see how this task will be rewarded.

If this task is selected for payment, you get **$0.50 per problem you solve correctly**. (Your payment does not decrease if you provide an incorrect answer to a problem). We refer to this as the *piece rate* payment.

PLEASE DO NOT TURN OVER THIS PAGE UNTIL YOU ARE TOLD TO DO SO.

Page 2, Task 1

ID: «AddressBlock»

Question										
1.	13	+	45	+	23	+	71	+	78	=
2.	64	+	24	+	79	+	24	+	51	=
3.	82	+	44	+	52	+	17	+	42	=
4.	72	+	83	+	24	+	63	+	74	=
5.	27	+	35	+	25	+	71	+	48	=
6.	54	+	32	+	62	+	16	+	13	=
7.	23	+	61	+	72	+	23	+	56	=
8.	24	+	25	+	16	+	46	+	88	=
9.	57	+	62	+	34	+	34	+	75	=
10.	11	+	87	+	35	+	19	+	53	=
11.	54	+	58	+	65	+	23	+	31	=
12.	58	+	33	+	42	+	52	+	92	=
13.	84	+	40	+	16	+	52	+	21	=
14.	30	+	43	+	78	+	80	+	35	=
15.	63	+	20	+	24	+	12	+	57	=
16.	47	+	73	+	28	+	54	+	24	=
17.	23	+	46	+	72	+	48	+	23	=
18.	42	+	11	+	43	+	62	+	51	=
19.	67	+	29	+	74	+	12	+	92	=
20.	19	+	23	+	66	+	34	+	10	=
21.	51	+	67	+	72	+	46	+	49	=
22.	95	+	34	+	45	+	93	+	21	=
23.	26	+	83	+	62	+	36	+	34	=
24.	46	+	24	+	92	+	23	+	57	=
25.	12	+	86	+	59	+	75	+	41	=
26.	47	+	41	+	12	+	25	+	27	=
27.	62	+	77	+	53	+	64	+	34	=
28.	48	+	95	+	23	+	26	+	27	=
29.	72	+	42	+	92	+	40	+	67	=
30.	26	+	32	+	55	+	74	+	83	=
31.	74	+	52	+	64	+	23	+	41	=
32.	26	+	51	+	71	+	63	+	25	=
33.	68	+	46	+	85	+	32	+	45	=
34.	52	+	15	+	26	+	46	+	54	=
35.	72	+	66	+	23	+	36	+	25	=
36.	95	+	73	+	82	+	45	+	14	=
37.	24	+	82	+	52	+	72	+	48	=
38.	44	+	45	+	14	+	37	+	23	=
39.	35	+	23	+	64	+	24	+	66	=
40.	14	+	61	+	34	+	87	+	88	=

ID: «AddressBlock»

How do you think your Task 1 score ranked within your group? (Please circle)

1st   2nd   3rd   4th

Task 2: Tournament

In this task, you will again be asked to calculate the sum of five randomly chosen two-digit numbers.

NOTE: You cannot use a calculator to determine this sum, however you may use the scrap paper provided at your desk. Use the pen at your desk to complete the task.

When we tell you to start, you have five minutes to complete as many questions as you can. At the end of the five minutes, we will collect your question sheets and mark them to determine your Task 2 score, which is the number of questions you answer correctly.

Please continue reading the instructions on this page on your own to see how this task will be rewarded.

If this task is selected for payment, you get **$1 per correct answer if you are a winner in your group and nothing if you are not a winner**. We refer to this as a *tournament* payment.

How do we determine who wins the tournament?
Winners are the two participants who give the highest number of correct answers in your group.

The following two examples illustrate some possible scenarios:

Example 1 – Suppose the scores are as follows:

Participant	Task 2 Score
A (man)	9
B (man)	8
C (woman)	3
D (woman)	2

The winners are A (man) and B (man).

Example 2 – Suppose the scores are as follows:

Participant	Task 2 Score
A (man)	3
B (man)	5
C (woman)	8
D (woman)	9

The winners are C (woman) and D (woman).

(If there is a tie, winners will be drawn randomly from the eligible participants. For example, if C in Example 1 above had the same score as B, ie. 8, we flip a coin to decide whether B or C would be a winner. A would still be a winner.)

PLEASE DO NOT TURN OVER THIS PAGE UNTIL YOU ARE TOLD TO DO SO.

Page 2, Task 2
ID: «AddressBlock»

Question											
1.	32	+	61	+	54	+	11	+	55	=	
2.	48	+	12	+	37	+	27	+	27	=	
3.	24	+	52	+	45	+	35	+	19	=	
4.	33	+	46	+	63	+	72	+	94	=	
5.	37	+	38	+	23	+	34	+	56	=	
6.	93	+	37	+	86	+	80	+	86	=	
7.	45	+	66	+	71	+	34	+	47	=	
8.	83	+	49	+	82	+	23	+	82	=	
9.	44	+	23	+	45	+	24	+	47	=	
10.	53	+	47	+	72	+	26	+	12	=	
11.	28	+	29	+	99	+	71	+	65	=	
12.	15	+	70	+	34	+	34	+	37	=	
13.	82	+	28	+	51	+	72	+	75	=	
14.	61	+	52	+	26	+	45	+	64	=	
15.	65	+	17	+	23	+	73	+	24	=	
16.	34	+	86	+	45	+	12	+	57	=	
17.	50	+	77	+	53	+	35	+	88	=	
18.	60	+	41	+	63	+	87	+	59	=	
19.	23	+	35	+	82	+	56	+	34	=	
20.	28	+	17	+	29	+	20	+	76	=	
21.	38	+	95	+	68	+	11	+	56	=	
22.	45	+	65	+	75	+	83	+	92	=	
23.	72	+	43	+	33	+	64	+	18	=	
24.	56	+	48	+	47	+	41	+	85	=	
25.	84	+	23	+	38	+	69	+	34	=	
26.	65	+	75	+	64	+	21	+	57	=	
27.	93	+	56	+	34	+	60	+	32	=	
28.	53	+	23	+	12	+	58	+	34	=	
29.	12	+	76	+	13	+	77	+	62	=	
30.	72	+	35	+	46	+	43	+	86	=	
31.	56	+	68	+	76	+	26	+	57	=	
32.	23	+	34	+	53	+	97	+	61	=	
33.	74	+	78	+	56	+	64	+	73	=	
34.	63	+	55	+	89	+	10	+	98	=	
35.	88	+	72	+	20	+	75	+	22	=	
36.	24	+	12	+	13	+	76	+	57	=	
37.	95	+	54	+	44	+	34	+	34	=	
38.	66	+	52	+	34	+	82	+	48	=	
39.	74	+	18	+	69	+	38	+	95	=	
40.	13	+	14	+	32	+	57	+	63	=	

ID: «AddressBlock»

Task 3:

In this task, you will **not** be asked to calculate the sum of five randomly chosen two-digit numbers. Instead, you will be asked to indicate which of the following options you would like to apply to your **Task 1** score:

Option A: Piece Rate. Receive $0.50 per correct answer for your performance in Task 1.

Option B: Tournament. Receive $1 per correct answer if you are a winner of the tournament, and nothing otherwise.

<u>Please continue reading the instructions on this page on your own to see how this task will be rewarded.</u>

How do we determine who wins the tournament?

Winners are the two participants who gave the highest number of correct answers in your group in Task 1.

The following two examples illustrate some possible scenarios:

Example 1 – Suppose the scores are as follows:

Participant	Task 1 Score
A (man)	9
B (man)	8
C (woman)	3
D (woman)	2

The winners are A (man) and B (man).

Example 2 – Suppose the scores are as follows:

Participant	Task 1 Score
A (man)	3
B (man)	5
C (woman)	8
D (woman)	9

The winners are C (woman) and D (woman).

If this task is chosen for payment, you will be paid according to your choice.

NOTE: your choice does not affect anyone else's payoff, and your payoff is also not affected by anyone else's choice. For example, if only Participant C from Example 1 above chooses the tournament (and A,B,D choose the piece rate), C would still not win the tournament.

<u>Please circle below which one you would like to apply to your Task 1 score:</u>

Option A	Option B
Piece Rate	Tournament

ID: «AddressBlock»

Task 4:

In this task, you will again be asked to calculate the sum of five randomly chosen two-digit numbers.

NOTE: You cannot use a calculator to determine this sum, however you may use the scrap paper provided at your desk. Use the pen at your desk to complete the task.

When we tell you to start, you have five minutes to complete as many questions as you can. At the end of the five minutes, we will collect your question sheets, **but we will not mark them to determine your Task 4 score**. Instead, **your group members will determine your Task 4 score** and you will determine the Task 4 score of your group members. We call this a peer-review process.

<u>How does the peer-review process work?</u>

We will collect your question sheets from Task 4 and redistribute them to each of your group members in turn, and each will mark them. This means that you will mark the sheets of your group members and they will mark yours. You will be given the correct answers to Task 4, and you will be asked to enter the scores for your group members on a reporting sheet. The next page contains an example of the peer-review process.

For example, suppose you are Participant A (man). Therefore, you will mark the sheets of B (man), C (woman) and D (woman) in your group. We will distribute the question sheets three times, and each time you will receive either B, C or D's question sheet. Suppose that the question sheets of B, C and D look like this:

Participant B (man)	Participant C (woman)	Participant D (woman)
ID: 2901	ID: 8502	ID: 3267
1. 10+10+10+10+10 = 50 2. 20+20+20+20+20 = 100 3. 30+30+30+30+30 = 150	1. 10+10+10+10+10 = 50 2. 20+20+20+20+20 = 100 3. 30+30+30+30+30 = 150	1. 10+10+10+10+10 = 50 2. 20+20+20+20+20 = 150 3. 30+30+30+30+30 = 200

You will be given an answer sheet with the correct answers, and a reporting sheet. These are printed back-to-back on the same sheet of paper. They look similar to this:

Task 4 Answer Sheet	Task 4 Reporting Sheet
1. 10+10+10+10+10 = 50 2. 20+20+20+20+20 = 100 3. 30+30+30+30+30 = 150	ID: ___ ID: ___ ID: ___ Score:

You (in this example, you are participant A) will enter the ID and score of B, C and D from your group on the reporting sheet. Please write on the reporting sheet only— **do not write (e.g. ticks and crosses) on other people's question sheets.** The others in your group will fill out their own reporting sheets. Note that you will not be able to see the reporting sheets of your group members, and they will not see your reporting sheet. Thus, your responses and those of your group members remain anonymous.

At the end of the peer-review process, Participant A's reporting sheet might look like this:

Task 4 Reporting Sheet			
	ID: 2901	ID: 8502	ID: 3267
Score:	0	3	2

As you can see in this example, the scores in A's reporting sheet do not correspond to the number of correct answers (ID 2901 has 3 correct answers, ID 8502 has 3 correct answers, and ID 3267 has 1 correct answer).

We will collect your reporting sheets only. **We will not collect the question sheets,** please leave them at your desk.

How are Task 4 scores determined?

We will use the scores on the reporting sheets to determine your Task 4 score. More specifically, of the three reports that we receive from your group members about you, **we will randomly draw one report to determine your Task 4 score.** Using the example above, if we randomly draw Participant A's report to determine B's score, then the Task 4 score will be 0, even though B actually gave 3 correct answers.

Thus, your entries and the entries of your group members in the reporting sheets count, regardless of whether they are correct or incorrect.

How are payments determined in Task 4?

If this task is selected for payment, you will receive $1 **multiplied by your Task 4 score** if you are a winner of the tournament, and nothing otherwise.

Please continue reading the instructions on this page on your own to see how tournament winners are determined.

How do we determine who wins the tournament?

Winners are the two participants who have the highest Task 4 scores.

The following two examples illustrate some possible scenarios:

Example 1 – Suppose the scores are as follows:

Participant	Task 4 Correct Answers	Task 4 Score (given by peers)
A (man)	9	9
B (man)	8	8
C (woman)	7	7
D (woman)	2	2

The winners are A (man) and B (man). A can receive $9 and B $8.

Example 2 – Suppose the scores are as follows:

Participant	Task 4 Correct Answers	Task 4 Score (given by peers)
A (man)	9	8
B (man)	8	9
C (woman)	7	1
D (woman)	2	2

The winners are A (man) and B (man). A can receive $8 and B $9.

(If there is a tie, winners will be drawn randomly from the eligible participants. For example, if C in Example 1 above had the same Task 4 score as B, ie. 8, we flip a coin to decide whether B or C would be a winner. A would still be a winner.)

PLEASE DO NOT TURN TO PAGE 4 UNTIL YOU ARE TOLD TO DO SO.

Question											
1.	82	+	13	+	82	+	19	+	73	=	
2.	36	+	92	+	47	+	35	+	76	=	
3.	13	+	56	+	45	+	83	+	44	=	
4.	48	+	94	+	96	+	58	+	81	=	
5.	34	+	29	+	23	+	23	+	22	=	
6.	85	+	34	+	44	+	84	+	47	=	
7.	45	+	94	+	68	+	79	+	34	=	
8.	65	+	44	+	30	+	24	+	45	=	
9.	48	+	17	+	24	+	34	+	63	=	
10.	23	+	54	+	83	+	72	+	95	=	
11.	11	+	38	+	82	+	68	+	52	=	
12.	28	+	46	+	36	+	80	+	66	=	
13.	34	+	99	+	84	+	65	+	31	=	
14.	87	+	23	+	62	+	23	+	82	=	
15.	65	+	56	+	64	+	58	+	52	=	
16.	46	+	17	+	28	+	25	+	42	=	
17.	93	+	12	+	23	+	55	+	66	=	
18.	61	+	95	+	67	+	72	+	83	=	
19.	32	+	42	+	27	+	22	+	75	=	
20.	29	+	45	+	51	+	75	+	39	=	
21.	35	+	77	+	18	+	88	+	28	=	
22.	73	+	32	+	35	+	43	+	78	=	
23.	56	+	82	+	92	+	23	+	22	=	
24.	69	+	45	+	56	+	78	+	70	=	
25.	12	+	12	+	26	+	98	+	68	=	
26.	48	+	94	+	83	+	68	+	29	=	
27.	13	+	64	+	53	+	36	+	16	=	
28.	50	+	56	+	66	+	41	+	65	=	
29.	52	+	72	+	80	+	81	+	42	=	
30.	74	+	20	+	15	+	34	+	75	=	
31.	29	+	11	+	33	+	75	+	80	=	
32.	26	+	97	+	57	+	64	+	54	=	
33.	31	+	46	+	45	+	14	+	76	=	
34.	42	+	32	+	19	+	19	+	38	=	
35.	82	+	77	+	68	+	34	+	93	=	
36.	58	+	59	+	84	+	66	+	67	=	
37.	32	+	38	+	25	+	34	+	17	=	
38.	87	+	36	+	72	+	12	+	87	=	
39.	61	+	94	+	84	+	96	+	66	=	
40.	30	+	67	+	95	+	35	+	65	=	

If Task 4 scores were based on the number of questions you answered correctly in Task 4, how do you think would rank within your group? (Please circle)

1st   2nd   3rd   4th

Task 5

Before you do the peer-review, there is a final Task 5. In this task, you will **not** be asked to calculate the sum of five randomly chosen two-digit numbers. Instead, you will be asked to indicate which of the following options you would like to apply to your **Task 4** score:

Option A: Piece Rate. Receive $0.50 multiplied by your Task 4 score.

Option B: Tournament. Receive $1 multiplied by your Task 4 score if you are a winner of the tournament, and nothing otherwise. This is the same as how payments in Task 4 will be determined.

Remember that Task 4 scores are solely determined by the peer-review. Thus, they do not necessarily reflect the number of correct answers given.

Your choice does not affect the outcome of Task 4. However, it determines your payment in Task 5. Thus, if Task 5 is chosen for payment, you will be paid according to your choice.

NOTE: your choice does not affect anyone else's payoff in Task 5, and your payoff in Task 5 is also not affected by anyone else's choice. For example, if Participant A has the lowest Task 4 score in the group and he is the only person in his group who chooses the tournament (and B,C, D choose the piece rate), A would still not win the tournament.

Please circle below which one you would like to apply to your Task 4 score:

Option A	Option B
Piece Rate	Tournament

Questionnaire
ID: «AddressBlock»

1. What is your age in years? _____

2. What are you currently studying?
   Degree/s:_____
   Faculty/ies:_____
   Major/s:_____

3. How many bathrooms are there in your parents' house? (Please write a number)____

4. Do you have a car on campus? (Please circle one) Yes/No

5. Are you currently employed in any full-time or part-time job? (Please circle one)
   Part-time   Full-time   I am currently not employed

6. How do you think your Task 4 score (based on one randomly-drawn report) ranked within your group?
   1st   2nd   3rd   4th

7. Did you expect your peers to **misreport** your performance in Task 4?
   Yes, in all cases   Yes, in most cases   Yes, in some cases   No, not at all

8. Did you expect your male or female peers to report your performance **less** accurately?
   Male   Female   No difference

9. How many other participants in this experiment do you know? (Please write a number)____

10. Do you think that under-represented groups in the workplace, such as women, should receive affirmative action (special treatment to increase their chances of getting a job)? (Please circle one) Yes/No

11. Would you feel that it was unfair if you missed out on a job because of an affirmative action policy? (Please circle one) Yes/No

PLEASE TURN OVER

12. Please circle the likelihood that you would engage in each of the following activities:

	Extremely unlikely				Extremely likely
Admitting that your tastes are different from those of your friends	1	2	3	4	5
Arguing with a friend who has a very different opinion on an issue	1	2	3	4	5
Asking your boss for a raise	1	2	3	4	5
Betting a day's income at the horse races	1	2	3	4	5
Buying an illegal drug for your own use	1	2	3	4	5
Chasing a tornado by car to take photos that you can sell to the press	1	2	3	4	5
Cheating a fair amount on your income tax	1	2	3	4	5
Cheating on an exam	1	2	3	4	5
Co-signing a new car loan for a friend	1	2	3	4	5
Dating someone that you are working with	1	2	3	4	5
Deciding to share an apartment with someone you don't know well	1	2	3	4	5
Disagreeing with your father on a major issue	1	2	3	4	5
Driving home after you had three drinks in the last two hours	1	2	3	4	5
Eating 'expired' food products that still 'look okay'	1	2	3	4	5
Exploring an unknown city or section of town.	1	2	3	4	5
Forging somebody's signature	1	2	3	4	5
Frequent binge drinking	1	2	3	4	5
Going camping in the wild	1	2	3	4	5
Going down a ski run that is too hard or closed	1	2	3	4	5
Going on a safari in Kenya	1	2	3	4	5
Going on a two-week vacation in a foreign country without booking accommodations ahead	1	2	3	4	5
Going whitewater rafting at high water in the spring	1	2	3	4	5
Ignoring some persistent physical pain by not going to the doctor	1	2	3	4	5
Illegally copying a piece of software	1	2	3	4	5
Taking a medical drug that has a high likelihood of negative side effects.	1	2	3	4	5
Traveling on a commercial airplane	1	2	3	4	5
Plagiarizing a term paper	1	2	3	4	5
Engaging in unprotected sex	1	2	3	4	5
Investing 10% of your annual income in a blue chip stock.	1	2	3	4	5
Investing 10% of your annual income in a very speculative stock	1	2	3	4	5
Investing 10% of your annual income in government bonds (treasury bills).	1	2	3	4	5
Investing in a business that has a good chance of failing	1	2	3	4	5
Lending a friend an amount of money equivalent to one month's income	1	2	3	4	5
Moving to a new city	1	2	3	4	5
Never using sunscreen when you sunbathe	1	2	3	4	5
Never wearing a seatbelt	1	2	3	4	5
Not having a smoke alarm in or outside of your bedroom	1	2	3	4	5
Openly disagreeing with your boss in front of your coworkers	1	2	3	4	5
Periodically engaging in a dangerous sport (e.g. mountain climbing or sky diving).	1	2	3	4	5
Regularly riding your bicycle without helmet	1	2	3	4	5
Shoplifting a small item (e.g. a lipstick or a pen)	1	2	3	4	5
Smoking a pack of cigarettes per day	1	2	3	4	5
Speaking your mind about an unpopular issue at a social occasion	1	2	3	4	5
Spending money impulsively without thinking about the consequences.	1	2	3	4	5
Stealing an additional TV cable connection	1	2	3	4	5
Taking a day's income to play the slot-machines at a casino	1	2	3	4	5
Taking a job where you get paid exclusively on a commission basis	1	2	3	4	5
Trying bungee jumping	1	2	3	4	5
Using office supplies for your personal business	1	2	3	4	5
Wearing unconventional clothes	1	2	3	4	5

Appendix 2: Instructions for Gender-Quota Treatment (where they differ to the "No Gender-Quota" treatment)

Page 1

ID: «AddressBlock»

How do you think your Task 1 score ranked within your group? (Please circle)

1st   2nd   3rd   4th

Task 2: Tournament

In this task, you will again be asked to calculate the sum of five randomly chosen two-digit numbers.

NOTE: You cannot use a calculator to determine this sum, however you may use the scrap paper provided at your desk. Use the pen at your desk to complete the task.

When we tell you to start, you have five minutes to complete as many questions as you can. At the end of the five minutes, we will collect your question sheets and mark them to determine your Task 2 score, which is the number of questions you answer correctly.

<u>Please continue reading the instructions on this page on your own to see how this task will be rewarded.</u>

If this task is selected for payment, you get **$1 per correct answer if you are a winner in your group and nothing if you are not a winner**. We refer to this as a *tournament* payment.

How do we determine who wins the tournament?

There are two winners in your group of four, and **at least** one of them will be a woman. The winners are:
1) the woman who gives the highest number of correct answers and
2) the best other participant, regardless of gender (that is, either a man or a woman).

Thus, in this tournament, **women are advantaged**.

The following two examples illustrate some possible scenarios:

Example 1 – Suppose the scores are as follows:

Participant	Task 2 Score
A (man)	9
B (man)	8
C (woman)	3
D (woman)	2

Example 2 – Suppose the scores are as follows:

Participant	Task 2 Score
A (man)	3
B (man)	5
C (woman)	8
D (woman)	9

The winners are C (woman) and A (man), **even though B (man) gave more correct answers than C (woman).**

The winners are C (woman) and D (woman).

(If there is a tie, winners will be drawn randomly from the eligible participants. For example, if A in Example 1 above had the same score as B, i.e. 8, we flip a coin to decide whether A or B would be a winner. C would still be a winner, since she is the highest-scoring woman.)

PLEASE DO NOT TURN OVER THIS PAGE UNTIL YOU ARE TOLD TO DO SO.

ID: «AddressBlock»

Task 3:

In this task, you will **not** be asked to calculate the sum of five randomly chosen two-digit numbers. Instead, you will be asked to indicate which of the following options you would like to apply to your **Task 1** score:

Option A: Piece Rate. Receive $0.50 per correct answer for your performance in Task 1.

Option B: Tournament. Receive $1 per correct answer if you are a winner of the tournament, and nothing otherwise.

<u>Please continue reading the instructions on this page on your own to see how this task will be rewarded.</u>

How do we determine who wins the tournament?

There are two winners in your group of four, and **at least** one of them will be a woman. The winners are:
1) the woman who gave the highest number of correct answers in Task 1, and
2) the best other participant, regardless of gender (that is, either a man or a woman).

Thus, in this tournament, **women are advantaged**.

The following two examples illustrate some possible scenarios:

Example 1 – Suppose the scores are as follows:

Participant	Task 1 Score
A (man)	9
B (man)	8
C (woman)	3
D (woman)	2

Example 2 – Suppose the scores are as follows:

Participant	Task 1 Score
A (man)	3
B (man)	5
C (woman)	8
D (woman)	9

The winners are C (woman) and A (man), **even though B (man) gave more correct answers than C (woman).**

The winners are C (woman) and D (woman).

If this task is chosen for payment, you will be paid according to your choice.

NOTE: your choice does not affect anyone else's payoff, and your payoff is also not affected by anyone else's choice. For example, if only Participant B from Example 1 above chooses the tournament (and A, C, D choose the piece rate), B would still not win the tournament.

<u>Please circle below which one you would like to apply to your Task 1 score:</u>

Option A	Option B
Piece Rate	Tournament

ID: «AddressBlock»

If Task 4 scores were based on the number of questions you answered correctly in Task 4, how do you think would rank within your group? (Please circle)

1st   2nd   3rd   4th

Task 5

Before you do the peer-review, there is a final Task 5. In this task, you will **not** be asked to calculate the sum of five randomly chosen two-digit numbers. Instead, you will be asked to indicate which of the following options you would like to apply to your **Task 4** score:

Option A: Piece Rate. Receive $0.50 multiplied by your Task 4 score.

Option B: Tournament. Receive $1 multiplied by your Task 4 score if you are a winner of the tournament, and nothing otherwise. This is the same as how payments in Task 4 will be determined.

Remember that Task 4 scores are solely determined by the peer-review. Thus, they do not necessarily reflect the number of correct answers given.

Your choice does not affect the outcome of Task 4. However, it determines your payment in Task 5. Thus, if Task 5 is chosen for payment, you will be paid according to your choice.

NOTE: your choice does not affect anyone else's payoff in Task 5, and your payoff in Task 5 is also not affected by anyone else's choice. For example, if Participant A has the lowest Task 4 score in the group and he is the only person in his group who chooses the tournament (and B,C, D choose the piece rate), A would still not win the tournament.

Please circle below which one you would like to apply to your Task 4 score:

Option A	Option B
Piece Rate	Tournament

Task 4

How are Task 4 scores determined?

We will use the scores on the reporting sheets to determine your Task 4 score. More specifically, of the

Appendix II

Table A1: Willingness to compete in tournaments – Logit Specification

	(1)	(2)	(3)	(4)	(5)	(6)
	------- All -------		--- Low-Ability ---		--- High-Ability ---	
A. Non-Sabotage Round						
Gender Quota	-0.266	-0.247	-0.241	-0.224	-0.254	-0.239
	(0.096)***	(0.098)**	(0.143)*	(0.148)	(0.134)*	(0.133)*
Female	-0.284	-0.254	-0.136	-0.128	-0.426	-0.408
	(0.096)***	(0.100)**	(0.144)	(0.148)	(0.131)***	(0.135)***
Gender Quota × Female	0.392	0.381	0.335	0.348	0.426	0.420
	(0.091)***	(0.094)***	(0.154)**	(0.153)**	(0.112)***	(0.113)***
Task 1 Score		0.015		0.031		0.028
		(0.008)**		(0.021)		(0.018)
Risk Score		0.002		0.003		0.000
		(0.002)		(0.002)		(0.002)
B. Sabotage-Round						
Gender Quota	0.079	0.062	0.132	0.066	0.065	0.044
	(0.099)	(0.105)	(0.146)	(0.162)	(0.140)	(0.150)
Female	-0.062	-0.066	0.005	-0.038	-0.079	-0.064
	(0.101)	(0.106)	(0.147)	(0.155)	(0.147)	(0.155)
Gender Quota × Female	0.022	0.068	0.033	0.112	-0.046	-0.007
	(0.142)	(0.147)	(0.203)	(0.220)	(0.207)	(0.212)
Task 1 Score		-0.003		-0.048		-0.006
		(0.008)		(0.027)*		(0.014)
Risk Score		0.005		0.004		0.006
		(0.002)***		(0.002)		(0.003)**
Difference across rounds in:						
Gender Quota × Female	-0.403	-0.363	-0.307	-0.258	-0.526	-0.510
	(0.122)***	(0.137)***	(0.186)*	(0.216)	(0.143)***	(0.153)***
Observations	200	200	103	103	97	97

Notes: Estimates are based on logit models. Low-ability individuals are those performing below the average score (i.e., 12) in Task 1. Robust standard errors clustered by participant reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2: Wage differences across treatment environments

	(1) Gender Wage Gap	(2) Average Wage
Quota Tournament	-3.675	1.712
	(1.574)**	(0.325)***
Standard Tournament	4.225	2.282
	(1.997)**	(0.421)***
Standard Tournament Choice	3.355	1.707
	(1.656)**	(0.529)***
Quota Tournament Choice	-2.915	0.622
	(1.340)**	(0.381)
Standard Peer Review	2.892	1.283
	(1.802)	(0.599)**
Quota Peer Review	-0.082	0.583
	(1.319)	(0.578)
Choice Peer Review	1.035	0.541
	(1.252)	(0.634)
Quota Choice Peer Review	0.755	0.021
	(1.203)	(0.584)
Constant (Piece Rate)	0.635	5.868
	(0.384)*	(0.159)***
Observations	250	250

Notes: Omitted category is piece rate. Each unit of observation is a group-task. Because there are 50 groups of participants and 5 tasks, the total number of observations is 250. Robust standard errors reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$