

How Competitiveness May Cause a Gender Wage Gap: Experimental Evidence*

Matthias Heinz^{†1}, Hans-Theo Normann^{‡2} and Holger A. Rau^{§3}

¹University of Cologne

²Duesseldorf Institute for Competition Economics (DICE)

³University of Göttingen

October 2015

Abstract

We show that choices in competitive behavior may entail a gender wage gap. In our experiments, employees first choose a remuneration scheme (competitive tournament vs. piece rate) and then conduct a real-effort task. Employers know the pie size the employee has generated, the remuneration scheme chosen, and the employee's gender. Employers then decide how the pie will be split, as in a dictator game. Whereas employers do not discriminate by gender when tournaments are chosen, they take substantially and significantly more from female employees who choose piece-rate remuneration. A discriminatory wage gap occurs which cannot be attributed to employees' performance.

JEL Classification: C91, J16, M52

Keywords: dictator game, discrimination, gender wage gap, laboratory experiment, real-effort task.

*We are grateful to Uri Gneezy, Muriel Niederle, Nikos Nikiforakis, Beatrice Pagel, Gerhard Riener, Eric Schniter and Matthias Sutter for helpful comments. We are indebted to audiences at ESA Zurich 2013, Bavarian Micro Day Nuremberg 2013, Royal Economic Society Manchester 2014, 7th M-BEES 2014 in Maastricht, NERD Workshop 2014 Nuremberg and seminar audiences at Chapman University and the University of Duesseldorf for helpful comments. Last but not least we owe a debt of gratitude to Lise Vesterlund for sharing the z-Tree code with us.

[†]Email: heinz@wiso.uni-koeln.de, telephone: +49 221 470 7263.

[‡]Email: normann@dice.hhu.de, telephone: +49 211 8115297.

[§]Corresponding author, Email: holger.rau@uni-goettingen.de, telephone: +49 551 39 22281.

1 Introduction

Experimental research has repeatedly documented that women and men differ in their inclination to compete. Niederle and Vesterlund (2007) show that women tend to shy away from competition: they choose a competitive environment less often than men, even though they perform the underlying task equally well (see also Niederle et al., 2012). Dohmen and Falk (2011) find in a remuneration-choice experiment that women prefer fixed wages to a higher extent than men.^{1, 2} In a field experiment with children, Gneezy and Rustichini (2004) find that competition enhances the performance of males only. While this is not the case in Niederle and Vesterlund (2007) in that women also increase their performance under competition, women nevertheless tend to avoid competition whereas men embrace it (see Croson and Gneezy, 2009, for a survey of this literature).

In light of these results, it has been suggested that differences in competitiveness may entail pay gaps (Sutter and Rützler, 2010; Dohmen and Falk, 2011; Flory et al., 2014). If men choose competitive (better paid) environments more often, a wage gap emerges even when women and men perform equally well. Likewise, shying away from competitive environments may lead to women missing opportunities for better paid jobs.³ Thus, the differences in the attitude toward competition may cause a gender wage gap.

At first sight, it may appear that such a wage gap due to differences in the attitude toward competition would not be discriminatory as it would be consistent with the *equal pay for equal work* principle. Generally, various explanations have been proposed for the gender wage gap (Blau and Kahn, 1992; 1994; Fryer and Levitt, 2010): performance-based explanations refer to differences in, for example, education, skills, and occupational choices, but the wage gap may also result from discrimination when equally qualified and equally performing men and women are paid differently. It seems that a gender pay gap due to differences in the inclination to compete would fall into the group of non-discriminatory explanations as it is performance based.

In this paper we argue that gender differences in the attitude toward competition may bring along an additional discriminatory effect. In the aforementioned experiments, the payments for the different remuneration schemes are exogenously fixed and do not allow for discrimination by gender. Within the remuneration schemes, there is equal pay for equal work by design. In

¹Gneezy et al. (2009) show in a field setting that women's competitiveness strongly depends on the nature of the societies investigated (matrilineal vs. patriarchal). See also Hoffman et al. (2011).

²The issue that women sometimes perform less well in competitive environments can be related to the remuneration choices. Gneezy et al. (2003) show that women may be less effective than men in competitive environments even if they perform equally well in non-competitive settings.

³Buser et al. (2014) show that boys more frequently choose academic tracks which are math- and science-intensive as compared to women. It turns out that differences in competitiveness account for a large portion of the gender differences in track choices.

settings relevant in the field, this may not always be the case as remuneration is often flexible (bonuses, negotiations, award of perks, etc.)⁴ and is determined by superiors with the discretion to decide. If so, the remuneration decisions itself and the interaction with gender may play a role. How do employers reward performance given the choice for either competitive or non-competitive remuneration by females and males? The novelty of our experiments is that we allow for an *adjusted* wage gap to emerge (whereas in existing experiments the adjusted wage gap is inevitably zero). When superiors use their discretion to decide on payments, discrimination beyond what can be justified based on performance may result.

The goal of our research can also be put this way. The previous experimental literature mostly focused on the supply side (Bohnet et al., 2012, is an exception, see below). These papers analyze whether men and women differ in their labor supply decisions by way of their attitudes toward competitive environments. We mainly look at the demand side—the employers—and at the interaction between the supply and the demand side. We investigate how employers react to tournament vs. piece-rate choices by women and men.

To investigate this research question, we use a laboratory experiment with more than 600 participants. As in previous experiments, employees' choices and performances determine the pie size: they first choose a remuneration scheme (competitive tournament vs. piece rate) and then they conduct a real-effort task. This part of the experiment is similar to Niederle and Vesterlund (2007). We add to this a dictator-game stage (for example, Hoffman et al., 1996; Eckel and Grossman, 1998)⁵ where the pie is generated by the employees (recipients) and where employers (dictators) decide how much to take from the pie.⁶ When deciding, employers know the pie size, the remuneration choice, and the employee's gender. While this design is perhaps extreme in that employers have 100 percent discretion, it is suitable for identifying discriminatory effects which previous experimental research could not detect. We also conduct a control treatment where the remuneration scheme is given by a random computer move.

Our findings are as follows. When the employees decide to work under piece rate, employers

⁴On gender effects regarding negotiations and wage expectations, see Bowles et al. (2008), Rigdon (2012), and Babcock and Laschever (2003) which are discussed below.

⁵Other dictator games include, for instance, Cherry et al. (2002), and Cappelen et al. (2007). Erkal et al. (2011) analyze a multi-player, real-effort dictator game. Brock et al. (2013) and Cappelen et al. (2013) investigate dictator decisions involving risks.

⁶Dictator game experiments with earned money (real-effort experiments) were suggested by Ruffle (1998) where the performance of recipients in a general-knowledge contest determined the total surplus. In Cherry et al. (2002) the real-effort task was done by the dictator. Oxoby and Spraggon (2008) analyze both dictator- and recipient-earned wealth, also using GMAT questions. Further, see Konow (2000) and Heinz et al. (2012) for real-effort dictator games.

take substantially and significantly more from female workers. By contrast, employers do not discriminate based on gender when tournaments are chosen. Neither do they discriminate against women when the payment scheme is imposed by a random computer move. We thus observe a discriminatory wage gap which cannot be attributed to performance.

2 Experimental Design

Our design combines the remuneration-choice experiments of Niederle and Vesterlund (2007) with a real-effort dictator game (Ruffle, 1998; Konow, 2000; Cherry et al., 2002). The first two stages are similar to “Task 3” of Niederle and Vesterlund (2007) where subjects first choose a remuneration scheme and afterwards take part in a real-effort task. We add a third stage where a dictator game is played: dictators (employers) decide how to split the pie, and the recipients (employees) determine the pie size in the mathematical real-effort task.⁷

The game is played only once, and the timing is as follows:

1. Employees (recipients) choose between tournament and piece-rate remuneration,
2. Employees conduct a mathematical real-effort task which determines the pie size,
3. Employers (dictators) decide how much to take from the pie, after learning the following information: the employee’s gender, the size of the pie generated, the employee’s remuneration choice, and (if tournament was chosen) whether the employee was a winner of the tournament.

We now explain these stages in detail.

The remuneration choice

The participants acting as employees were told upfront that they would be partitioned into groups of six people and that all groups would consist of three female and three male participants. This was the only time that gender was mentioned in the experimental instructions.

Employees could choose between two remuneration schemes: tournament and piece rate. This part is similar to Niederle and Vesterlund’s (2007) “Task 3 — Choice of Compensation Scheme.”⁸

⁷This stage is somewhat similar to Bosman and van Winden’s (2002) power-to-take game where two players each earn an income in an individual real-effort task. One player (the authority) decides how much she will take away from the responder. In a second stage, the responder can punish the authority by destroying her own income. The transfer from the responder to the take authority is based on the the responder’s remaining income after the second stage.

⁸One difference to Niederle and Vesterlund’s (2007) design is that their participants do the real-effort task under

If tournament is chosen, payments depend on own performance and on the performance of other participants. Employees who choose tournament are compared to the other participants in their group, regardless of whether the other participants chose piece rate or tournament. When a subject achieves a score which is among the two best results in the group of six, she earns 90 “Talers” for each correct answer. Those participants who opted for tournament but did not manage to achieve one of the two best scores only receive 30 Talers for each correct answer.

For employees who choose piece rate, the payment rate for each correct answer is constant. They are given 50 Talers for each question that is answered correctly. Note that, for a one-third chance of winning the tournament, the expected payoff is the same under both payment schemes. We applied an exchange rate where 1.5 Taler corresponded to 1 cent.

The real-effort task

The real-effort task we use is the same as in Niederle and Vesterlund (2007). Employees have to add up five two-digit numbers. They are allowed to use scratch paper but pocket calculators are prohibited. The numbers are randomly drawn by the computer. Employees have 15 minutes in total for the real-effort task and can proceed at their own pace.

The real-effort task is as presented in Table 1. Subjects have to fill the correct answer into the blank box. After submitting an answer, subjects are immediately informed whether it was correct. On the screen, there is an overview of the number of correct and wrong answers.

58	83	76	13	85	
----	----	----	----	----	--

Table 1: Example of a problem in the real-effort task

The employer decision

Employers have complete discretion as to how to divide the pie generated by the employees (Ruffle, 1998). In the instructions, the employer decision reads as follows: “You decide on the allocation of the money between you and the assigned other participant.”

both the non-competitive piece-rate condition and the competitive tournament environment before deciding which payment scheme to take. In our setup, the remuneration choice is upfront; our subjects knew the task to be performed, but had no experience as to how well they would do. Another difference to Niederle-Vesterlund (2007) is that our subjects had 15 minutes rather than 5 to solve mathematical problems. Finally, and perhaps most importantly, there is no dictator-game stage in Niederle and Vesterlund (2007). Our subjects knew that the final payoffs would be determined by the employer.

Before deciding, employers are informed of the size of the pie, the employee’s remuneration choice and the employee’s gender. Employers are also informed of whether the employee is a winner of the tournament, provided tournament is chosen. Table 2 summarizes the feedback given to employers. Employees have to fill in this form themselves, but the entries are checked by the experimenters.

Participant ID	# 01
Sex	female
Chosen remuneration scheme	tournament
Among the two best performers (if tournament)	yes
Earned sum of money	17.50€

Table 2: Example of the feedback given before employers decide on the final allocation

Treatments

We have three treatments, called *Choice*, *Feedback*, and *Random*. The main treatment, *Choice*, is as in the sections above. It includes the three aforementioned stages where, in stage 1, the remuneration scheme is chosen by the employees.

In *Feedback* we varied the information feedback. In addition to the items listed in Table 2, we gave information on the relative performance of the employee. For each possible outcome of the remuneration choice (piece rate, tournament winner, tournament loser), we told the employers the corresponding average earnings in *Choice*.⁹

In *Random* there is no choice at stage 1; instead, the remuneration scheme is given by a random computer draw. To ensure comparability of treatments, we explained both payment schemes to employers and employees before they were told the relevant scheme that had been allocated to them. Participants were told that a random draw would select whether an employee would work under piece-rate or tournament conditions. This random draw was designed such that it roughly matched the frequency of tournament choices we observed in the *Choice* and *Feedback* treatments (52%)

⁹We told the employers that “a similar experiment has already been conducted some time ago. The earned sum of money of participants A who had chosen piece rate was 9.60 euros.” In cases where the employee chose tournament and won [lost] we wrote “tournament and who were [not] among the two best performers” instead of “piece rate,” and the amounts were 26.20 and [5.80] euros, respectively. Employees did not know that the employers would be informed of the average earnings when they chose the remuneration scheme and worked on the real-effort task.

which we rounded down to 50 percent, so precisely half of the employees worked under tournament conditions. The groups of six participants were randomly composed, so they potentially contained both piece-rate and tournament employees.

3 Insights from previous studies

A first finding in the literature is that employers reward good employee performance. This has been shown in dictator games where, as in our experiment, the recipients earn the surplus in real-effort tasks (Ruffle, 1998; Oxoby and Spraggon, 2008; Heinz et al., 2012). Depending on their performance, recipients either generate a small or a large pie. The surplus generated is hence a clear-cut measure of relative performance which the dictators may take into account. Typically, dictators give more to well-performing recipients; low-performing recipients receive less on average and are frequently given zero.¹⁰

A second group of findings concerns the accountability (Konow, 2000) of decisions. Accountability means that people will be called to account only for discretionary decisions, but not for exogenous factors which they cannot reasonably influence.¹¹ Konow (2000) analyzes real-effort dictator games and shows that dictators' decision making will be affected only when the recipient is accountable for certain variables.

What do these insights imply for our setup? As far as performance is concerned, note that we give feedback on relative performance only in *Feedback* whereas performance is difficult to assess in *Choice* and *Random*. In all our treatments, employees' absolute performance is observable for the employers only in terms of money earned. From this, they can also conclude the number of correctly solved questions. In *Choice* and *Random*, however, it is difficult to assess how good a single employee's performance is. Due to this ambiguity, we expect that employers will not base their decisions on performance only but also on the information we provide. For example, employers may believe that employees work harder under tournament conditions and may thus reward employees better in that case.¹² In *Feedback*, we provide an unambiguous feedback on relative recipient performance, so we expect employers to reward good performance, as in previous studies. Regarding the accountability argument, participants are fully responsible for their decisions in

¹⁰In Heinz et al. (2012) this result is driven by female dictators being more generous than male dictators.

¹¹Cappelen et al. (2007) suggest liberal egalitarianism as an equivalent concept: that people should only be held responsible for their *choices* (for example, effort). This is in contrast to libertarian notions which also hold people responsible for their ability and talent on the one hand, and to notions of strict distributional egalitarianism on the other.

¹²In these treatments, we realistically capture situations in the field: If performance and success were always fully obvious to employers, performance evaluation and fair remuneration would perhaps be less of a problem in the field.

Choice and *Feedback* whereas they are not accountable for it in *Random*.

Putting these arguments together, we state our main hypothesis as follows. In *Choice*, employers will reward employees who choose tournament more generously than those who choose piece-rate remuneration. This will not occur in the other treatments because, in *Feedback*, employers have direct evidence on the relative performance and, in *Random*, subjects cannot be held accountable for the remuneration choices.

The existing literature also suggests that employers may discriminate based on employee gender. One hypothesis is that an employer's assessment of a performance is affected by a gender stereotype. It is well documented that a significant fraction of the population shares a stereotypical belief that men are superior at mathematical tasks.¹³ Recently, Bohnet et al. (2012) reported an experiment that supports this hypothesis. Using Niederle and Vesterlund's (2007) mathematical task, they let employers decide whether to hire a job candidate. The information they give to employers is the candidate's gender and his or her past performance in the task. For this mathematical task, the authors find that employers are more likely to employ a male employee, suggesting that they believe the male candidate will perform better in a (payoff-relevant) second round.¹⁴ In our experiment, biased employers might take the information "male employee" as an indicator of good performance and pay them better, accordingly.

Another hypothesis is based on findings that male employees expect and demand more in negotiations.^{15, 16} For example, Bowles et al. (2008) report results from negotiations where men penalize female job candidates more than male candidates for initiating negotiations. Women are therefore less inclined to initiate negotiations with male evaluators. In a setting without information on employees' gender, Rigdon (2012) finds that women request lower wages than men and, in turn, female negotiators receive less. She shows that this effect vanishes when female negotiators are given information in advance on the average requests of men. Babcock et al. (2003) report similar results. In our setting, when employers expect male employees to demand more, they might be inclined to give more to men in order to meet this expectation.

¹³See, for instance, Fennema and Sherman (1977, 1978), Eccles et al. (1990), Swim (1994), Spencer et al. (1999).

¹⁴The main finding of Bohnet et al. (2012) is that this discriminatory effect vanishes when the performance of one male and one female candidate are simultaneously presented to employers. They also show that the stereotype effect does not occur when they employ a non-mathematical task.

¹⁵We are grateful to Muriel Niederle for suggesting this hypothesis to us.

¹⁶Eckel and Grossman (2001) also report this finding in the ultimatum game, that is, they find male second-movers to be more demanding than female responders.

4 Procedures

Before entering the laboratory, participants were randomly assigned to two groups of equal size and divided into two separate rooms (A and B). Upon arrival, a random draw for each individual participant determined the A vs. B allocation. Subjects in room A were the employees. Subjects in room B were the employers. We ensured that half of the participants in room A were female.

The employees in room A were informed that they had been partitioned into cohorts of six people and that all groups consisted of three female and three male participants. (The instructions said: “In your group there are exactly three women and three men.”) In the meantime, the employers in room B had to wait approximately 20 minutes. They knew the nature of the task the employees had to conduct, and they also knew about the remuneration choice. Employers were not allowed to talk during the waiting time. After the employees had completed the real-effort task, we informed them about how much money they had generated. Subjects who had chosen the tournament were also informed as to whether or not they had won the tournament.

Next, employees were given a form and asked to fill in their participation number, gender, the chosen payment scheme and, in the case of tournament remuneration, whether they were one of the two best participants in their group. Finally, they had to fill in the amount of money they had generated. The experimenters checked that all forms had been filled in correctly. Each sheet was collected and put in an envelope. The envelopes were brought to the employers who each received one envelope. The allocation was random and the randomization was publicly visible. Employers had to decide in a dictator-like fashion on the allocation of money. Finally, we informed the participants in room A of the allocation decision of their employer. Subjects in rooms A and B were paid privately. All participants read the same instructions. They were fully informed about all details of the procedures of the experiment at the beginning. We applied neutrally framed treatments without using the words employer, employee, worker. Instead, the instructions said “participant A/B.”

The experiment was conducted in the FLEX laboratory at the Goethe-University Frankfurt. The subject pool consisted of 632 students from various fields. We conducted four sessions of each treatment, each session with 48 (one session: 36) subjects. We had 192 subjects participating in *Choice*, 192 in *Random*, and 180 in *Feedback*. Three employee-employer pairs were discarded from the data set.¹⁷ After dropping these six subjects from the data, we were left with 558

¹⁷One participant in the *Choice* treatment did not give the information about gender (this employee wrote “irrelevant” in the form) and we decided not to insist on the correction of the entry. (We know that this subject was female, however, we still cannot include the data because the employer did not know the employee’s gender when deciding.) One male employee in the *Choice* treatment appears to be an outlier: he solved 108 problems, 106 of them correctly. This is more than three times the average number of correct scores and still far above the second-highest

participants whose decisions we could use: There were 94 employee-employer pairs in *Choice* (47 female employees), 95 in *Random* (47 female employees), and 90 in *Feedback* (45 female employees). The remaining 68 subjects were employed in three additional sessions which were designed to elicit employee expectations. See section 6 below.

One session typically lasted about 75 minutes, subjects earned on average 17.26 euro including a 5 euro show-up fee. Subjects were recruited with ORSEE (Greiner, 2004). The experiment was conducted using z-Tree (Fischbacher, 2007) with an adapted version of the code used in Niederle and Vesterlund (2007).

5 Results

Throughout this section we employ the following variables. The *pie size* is the amount of money employees generate by solving the problems. We denote the absolute amount of money employers take from the pie and give to the employees by *employee income*. This is the employers' one and only decision. The *employee share* measures the employees' income relative to the pie generated. We also discuss employees' *performance* which measures the number of correctly solved problems. We will present the results in a sequence that reverses the order of decision making. We start with the employers' decisions and move backwards to employees' performance and remuneration decisions.

5.1 Employee share

Table 3 summarizes the employer decisions. Our research question is how the choice of tournament vs. piece-rate remuneration affects the share of the pie the employers give away. The main variable of interest is thus *employee share*. It is suitable to identify discrimination since it measures the employees' income relative to performance.

scorer (71 questions, 63 correct). Furthermore, one male employer of the *Random* treatment indicated after the experiment that he had misinterpreted the value given and requested permission to change his giving rate. Our results do not change qualitatively when we include these observations in the analysis.

	piece rate			tournament		
	men	women	avg.	men	women	avg.
<i>Choice</i> ($n = 94$)						
employee share	0.36 (0.22)	0.18 (0.21)	0.26 (0.23)	0.31 (0.23)	0.30 (0.19)	0.31 (0.21)
employee income (euro)	3.09 (2.31)	2.03 (2.50)	2.53 (2.45)	5.71 (6.70)	3.74 (3.78)	4.79 (5.56)
pie size (euro)	8.90 (2.90)	10.16 (2.83)	9.57 (2.90)	16.77 (12.58)	13.59 (9.09)	15.29 (11.09)
<i>Feedback</i> ($n = 90$)						
employee share	0.22 (0.22)	0.32 (0.22)	0.29 (0.22)	0.35 (0.21)	0.23 (0.21)	0.31 (0.21)
employee income (euro)	2.65 (3.19)	3.47 (3.06)	3.27 (3.07)	6.04 (5.47)	5.29 (6.03)	5.80 (5.61)
pie size (euro)	10.52 (4.30)	10.54 (3.90)	10.54 (3.94)	16.38 (10.44)	18.48 (13.04)	17.05 (11.25)
<i>Random</i> ($n = 95$)						
employee share	0.24 (0.25)	0.32 (0.22)	0.28 (0.24)	0.33 (0.19)	0.33 (0.23)	0.33 (0.21)
employee income (euro)	2.80 (3.22)	3.21 (2.38)	3.00 (2.81)	5.68 (5.78)	3.88 (4.86)	4.78 (5.36)
pie size (euro)	10.67 (3.34)	10.01 (2.91)	10.35 (3.12)	17.27 (12.56)	9.68 (8.05)	13.48 (11.12)

Table 3: Employers decide on *employee income*, taken as given the *pie size* and the remuneration scheme. The *employee share* is the ratio of *employee income* and *pie size*. Standard deviations in parentheses.

A conspicuous finding in Table 3 occurs in *Choice* when piece rate is chosen. Employers pay a much smaller fraction to female employees (18%) than to men (36%). Indeed, the *employee share* for women is only half of the share men receive. This difference is significant (Mann-Whitney test, $p = 0.005$). The result still holds after we Bonferroni-correct the significance level to $\alpha = 0.1/6 = 0.017$ because of multiple comparisons.¹⁸

We highlight this finding in Figure 1 which shows the *employee share* for women and men in *Choice*. We find in the *Choice* treatment that the *employee share* between men and women under piece rate is also in great contrast to the other five gender differences in Table 3, which are all insignificant. Specifically, in *Choice* under the tournament payment scheme, no gender differences whatsoever occur: employers give 31 percent of the pie to male employees and 30 percent to female employees who opted for the tournament remuneration.

¹⁸With three treatments and two remuneration schemes, we have $n = 6$ male-female comparisons of *employee share* in Table 3.

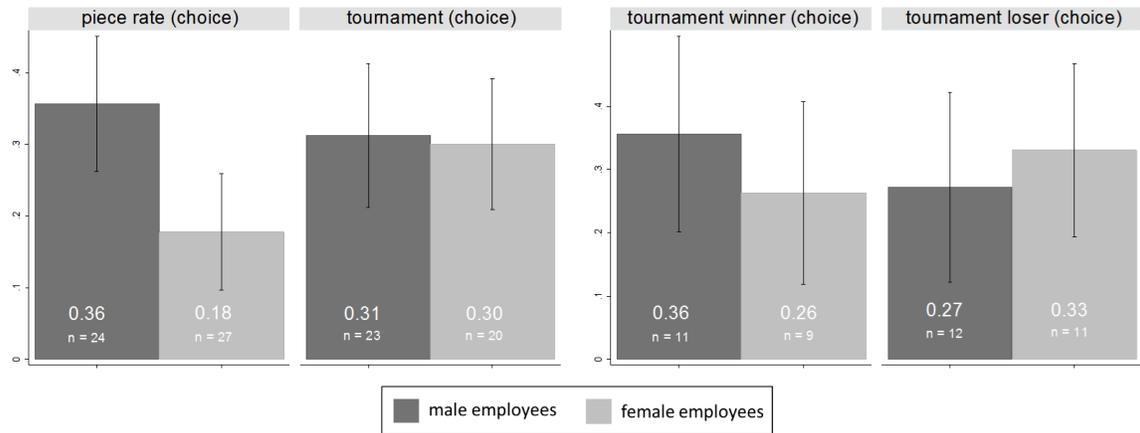


Figure 1: Columns show the average *employee share*, with error bars, for women and men in treatment *Choice*.

Splitting the tournament data further into winners and losers does not lead to unambiguous insights, see Figure 1. Male winners of the tournament have an average *employee share* larger than losers, whereas for women it is the other way round. These differences are not significant (Mann-Whitney tests, $p = 0.382$ (winners); $p = 0.554$ (losers)).

Further entries in Table 3 help to explain the gender gap of *employee share* in *Choice*. Under piece rate, women actually generate a larger *pie size* than men. Nevertheless, women receive a lower absolute income than men. Altogether, receiving a smaller absolute income out of a larger pie implies a drastically reduced *employee share*. Intriguingly, male employees who won the tournament never receive a payment of zero.

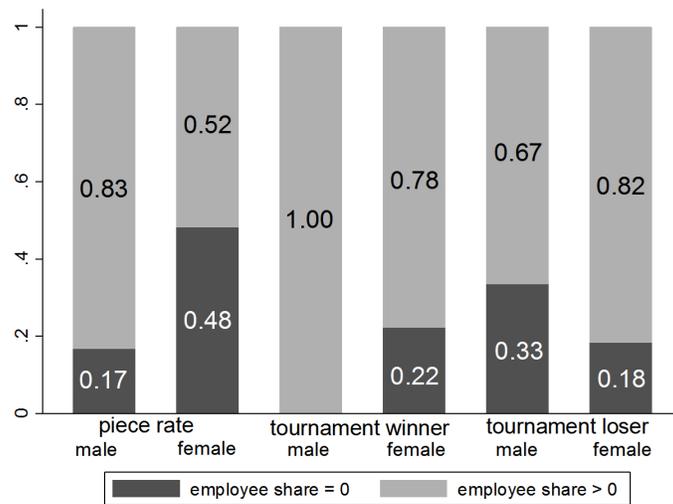


Figure 2: Distributions of giving rates conditional on remuneration choices and gender in *Choice*

Looking at the distribution of *employee share* is helpful. We categorize here into *employee share* = 0 and *employee share* > 0. Figure 2 shows the resulting distributions for the *Choice* treatment.¹⁹ The case where nothing is given most often occurs for female employees under piece rate (48%).

We investigate the significance of these effects using probit regressions. Table 4 analyzes for *Choice* the cases where employees received a share of zero. Model (1) includes dummy variables controlling for employees' gender and the remuneration conditions. In Model (2) we interact the remuneration conditions with employees' gender. In regression (1) we find that female employees are significantly more likely to receive a zero payment in *Choice*. Regression (2) confirms the finding of Figure 2, showing that the former result is exclusively driven by female employees who chose piece rate.

	employee share = 0	
	(1)	(2)
<i>female employee</i>	0.602** (0.294)	-0.478 (0.578)
<i>piece rate</i>	0.156 (0.337)	-0.537 (0.482)
<i>tournament winner</i>	-0.710 (0.491)	-4.917 (263.150)
<i>piece rate</i> × <i>female employee</i>		1.399** (0.696)
<i>tournament winner</i> × <i>female employee</i>		5.061 (263.151)
<i>constant</i>	-0.931*** (0.317)	0.102 (-0.431)
<i>pseudo R</i> ²	0.081	0.134
<i>observations</i>	94	94

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Table 4: Probit regressions of the cases where employees obtained a share of zero in *Choice*.

Does employees' performance matter for the *employee share*? Figure 3 presents scatter plots showing the relation between employee performance and *employee share* in our three treatments. In *Choice* (left panel) it can be seen that there exists no significant correlation between employees' performance and the employee share. At the same time, the scatter plots show that employees obtain a higher *employee share* when performing better in *Feedback* and *Random*. This is a first hint that employers concentrate on factors other than performance when deciding on employees' remuneration in the *Choice* treatment.

¹⁹See figures 4 and 5 in Appendix A for presentations of the *Feedback* and *Random* treatments.

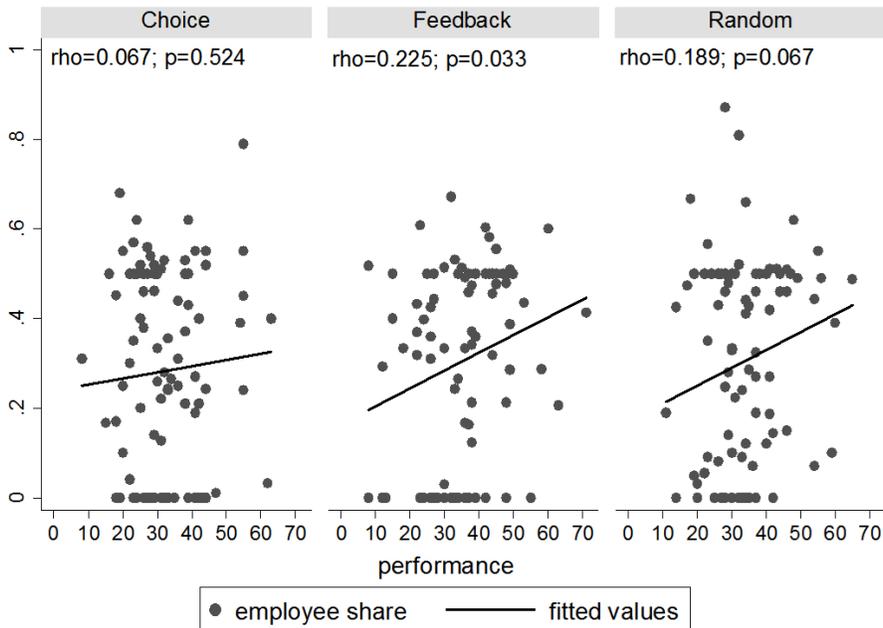


Figure 3: The relation between employees performance and the *employee share* they obtained.

We now provide comprehensive regression analyses of *employee share* for a full understanding. We use tobit regressions because the variables are left censored and, indeed, have many observations with a zero *employee share*. The omitted treatment, captured by the constant, is *Feedback*. Consider the three regressions reported in Table 5 in turn. Regression (1) uses exogenous regressors only (that is, dummies for the treatments, employer and employee gender). We see that *female employers* give significantly more. This finding also holds in the other regressions. Neither our treatment variables nor the employee gender turn out to be significant in regression (1). Regression (2) adds the performance-related variables *performance*, *piece rate*, and *tournament winner*. Again, we observe no significant regressors except *female employers*. Tournament winners do not receive a higher *employee share*. Finally, regression (3) adds all relevant interactions. Those with *Choice* can be found in the middle of the table and those with *Random* at the bottom.

As in Table 3, the low employee share given to women in *Choice* under piece-rate remuneration stands out in (3): $piece\ rate \times female\ employee \times choice$ is economically and statistically significant. This strong effect is only partially compensated by women's significantly higher *employee share* with piece-rate remuneration ($piece\ rate \times female\ employee$) as the sum of both effects is negative (Wald test, $p = 0.004$). Also (borderline) significant is $piece\ rate \times choice$, but this effect applies to both sexes and thus does not change the gender gap. To sum up, the regression analysis confirms the descriptive analysis above: a significant gender gap in terms of *employee share* occurs only in *Choice* under piece-rate conditions. No other significant effects occur.

	employee share		
	(1)	(2)	(3)
<i>choice</i>	-0.016 (0.042)	-0.002 (0.042)	0.026 (0.164)
<i>random</i>	0.020 (0.041)	0.028 (0.041)	-0.017 (0.161)
<i>female employee</i>	-0.041 (0.034)	-0.030 (0.034)	-0.245* (0.129)
<i>female employer</i>	0.094*** (0.034)	0.087** (0.334)	0.076** (0.334)
<i>piece rate</i>		-0.009 (0.040)	-0.129 (0.113)
<i>tournament winner</i>		0.022 (0.054)	0.006 (0.094)
<i>performance</i>		0.003* (0.002)	0.004 (0.003)
<i>piece rate</i> × <i>female employee</i>			0.360** (0.166)
<i>tournament winner</i> × <i>female employee</i>			0.176 (0.168)
[Interactions with <i>choice</i>]			
<i>piece rate</i> × <i>choice</i>			0.246* (0.147)
<i>tournament winner</i> × <i>choice</i>			0.083 (0.158)
<i>female employee</i> × <i>choice</i>			0.330* (0.170)
<i>piece rate</i> × <i>female employee</i> × <i>choice</i>			-0.680*** (0.215)
<i>tournament winner</i> × <i>female employee</i> × <i>choice</i>			-0.361 (0.235)
<i>choice</i> × <i>performance</i>			-0.003 (0.004)
[Interactions with <i>random</i>]			
<i>piece rate</i> × <i>random</i>			-0.004 (0.147)
<i>tournament winner</i> × <i>random</i>			-0.101 (0.155)
<i>female employee</i> × <i>random</i>			0.166 (0.161)
<i>piece rate</i> × <i>female employee</i> × <i>random</i>			-0.153 (0.208)
<i>tournament winner</i> × <i>female employee</i> × <i>random</i>			0.115 (0.246)
<i>random</i> × <i>performance</i>			0.002 (0.004)
<i>constant</i>	0.227*** (0.040)	0.102 (0.072)	0.118 (0.109)
<i>pseudo R</i> ²	0.049	0.086	0.197
<i>observations</i>	279	279	279

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Tobit regression of the share paid to the employees by employers.

5.2 Employee income

We now turn to the absolute income employees receive (*employee income*). This is the actual decision the employers make. Corresponding to the low *employee income* in Table 3 for women who choose piece rate in *Choice*, women also get a 34 percent lower *employee income* in that case (2.03 vs. 3.09 euros). This difference fails to be significant when we Bonferroni-correct for six multiple comparisons (Mann-Whitney, test $p = 0.077$). The other five Mann-Whitney tests comparing female-male *employee income* are all $p > 0.238$.

Table 6 reports regressions on *employee income* with the same regressors as in Table 5. In regression (1), female employees obtain an income that is approximately 1.4 euros lower than men. But this pay gap disappears when we control for *performance* in regressions (2) and (3). Female employees give more income to employees in all regressions. In (2) and (3), *tournament winner* and *performance* are highly significant. The variable *performance* strongly correlates to *pie size* and thus has a major impact on *employee income*. Regression (3) includes the interactions, whereas regression (4) uses a different dependent variable. Here, we normalize the employee income by performance. That is, in (4) we analyze the euro payment for each correctly solved problem. All regressors are the same as in regression (3), except that we exclude the *performance* regressors. This approach accounts for the difference between the payments to a woman and a man, given that they have the same performance in the task.²⁰ In models (3) and (4) we confirm that female employees receive a significantly lower income than men when choosing the piece-rate remuneration. However, the effect becomes borderline insignificant in model (3) when we add *piece rate* \times *female employee* (Wald test, $p = 0.111$). Whereas, the effect remains significant in model (4) (Wald test, $p = 0.016$).²¹ Our findings can be summarized as follows:

Result 1. *Comparing payments to women and men, we find that women earn less both in relative terms (employee share) and absolute terms (employee income) for piece-rate choices in treatment Choice. This effect is significant in a multiple regression analysis for employee share and is borderline insignificant for employee income. The gender differences in the other treatments and remuneration schemes are insignificant.*

²⁰We thank an anonymous referee for raising this point.

²¹An unexpected finding is that *tournament winner* \times *female employee* \times *random* is borderline significant in Model (3). However, the result should be treated with caution, as we only have four observations where women won the tournament in *Random*.

	employee income			normalized
	(1)	(2)	(3)	(4)
<i>choice</i>	-1.171 (0.835)	-0.195 (0.592)	1.662 (2.360)	-0.015 (0.038)
<i>random</i>	-0.478 (0.828)	0.386 (0.589)	0.081 (2.322)	0.006 (0.037)
<i>female employee</i>	-1.421** (0.674)	-0.451 (0.482)	-1.550 (1.887)	-0.072 (0.048)
<i>female employer</i>	1.888*** (0.683)	1.144** (0.481)	1.008** (0.480)	0.028** (0.012)
<i>piece rate</i>		1.458** (0.628)	0.216 (1.633)	0.004 (0.042)
<i>tournament winner</i>		8.183*** (0.735)	5.745*** (1.348)	0.165*** (0.033)
<i>performance</i>		0.171*** (0.025)	0.190*** (0.040)	
<i>piece rate</i> × <i>female employee</i>			2.648 (2.418)	0.111* (0.061)
<i>tournament winner</i> × <i>female employee</i>			-0.112 (2.418)	0.037 (0.061)
[Interactions with <i>choice</i>]				
<i>piece rate</i> × <i>choice</i>			2.080 (2.129)	0.075 (0.054)
<i>tournament winner</i> × <i>choice</i>			0.992 (2.252)	0.009 (0.052)
<i>female employee</i> × <i>choice</i>			2.227 (2.473)	0.095 (0.063)
<i>piece rate</i> × <i>female employee</i> × <i>choice</i>			-5.709* (3.112)	-0.213*** (0.079)
<i>tournament winner</i> × <i>female employee</i> × <i>choice</i>			-3.627 (3.367)	0.121 (0.085)
<i>choice</i> × <i>performance</i>			-0.064 (0.062)	
[Interactions with <i>random</i>]				
<i>piece rate</i> × <i>random</i>			0.000 (2.125)	0.005 (0.054)
<i>tournament winner</i> × <i>random</i>			-1.375 (2.213)	-0.025 (0.051)
<i>female employee</i> × <i>random</i>			0.711 (2.357)	0.057 (0.060)
<i>piece rate</i> × <i>female employee</i> × <i>random</i>			-0.581 (3.015)	-0.055 (0.077)
<i>tournament winner</i> × <i>female employee</i> × <i>random</i>			5.914* (3.522)	0.093 (0.090)
<i>random</i> × <i>performance</i>			0.112 (0.061)	
<i>constant</i>	3.452*** (0.010)	-4.653*** (1.029)	-5.114*** (1.569)	0.032 (0.025)
<i>pseudo R</i> ²	0.049	0.115	0.129	-0.720
<i>observations</i>	279	279	279	279

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Tobit regression of the euros paid to the employees by employers. Regression (4) analyzes the euro payment for each correctly solved problem.

Our findings on tournament winners and losers appear unexpected at first sight. Table 5 suggests that winners of the tournament do not receive a larger *employee share* than losers. But should the superior performance of the winners not be much better rewarded by employers? Looking at the absolute *employee income* solves the puzzle and, indeed, *tournament winner* is highly significant in regressions (2) and (3) of Table 6. Tournament winners generate a much larger pie (because they perform better and obtain the large payment of 90 Taler for each correctly solved problem): on average, winners generate a *pie size* of 26.5 euros. By contrast, the average *pie size* for tournament losers is 5.5 euros. Now consider the overall average *employee share* of 30 percent. This would amount to an *employee income* of roughly 8 euros for a tournament winner (one could consider this a reasonable sum for a 75-minute experiment) but only mere 1.65 euros (by most standards a petty amount) for a tournament loser. So, seen from the perspective of the absolute payments, winners and losers are treated rather differently but the *employee share* would nevertheless be the same.

5.3 Employer gender

Both tables 5 and 6 strongly suggest that female employers give more. Looking at simple non-conditional averages, we find that female employers give 32.7 percent of the pie whereas male employers give 25.9 percent. In terms of *employee income*, women give 1.26 euros more than men. Both differences are significant (Mann-Whitney test, $p = 0.012$) and confirm the findings in Heinz et al. (2012). We summarize:

Result 2. *Female employers give significantly more both in absolute terms (employee income) and relative terms (employee share).*

More pertinent to our research is the question of to what extent the gender of the employer matters regarding our main research question. The problem is that if we condition employer behavior on employee gender and payment scheme, we would need to split our data set into eight different subcases in each treatment. So we only have a few observations in each category and the findings are somewhat speculative in nature. Nevertheless, we report on the most conspicuous results. Consider the cases where piece-rate remuneration is chosen in *Choice*. Both male and female employers give 14–15 percentage points more to male employees in this case, although women do so at a higher level (men give 27 percent to male employees and 12 percent to female workers; female employers allocate 43 percent to male workers and 29 percent to women). Worth pointing out is the rather low share of 12 percent that male employers pay to female employees. Instances with a zero giving rate are also telling: when men pay women under piece-rate, 11 out

of 18 employers gave zero and one more employer gave next to zero. By contrast, male employers pay significantly more (27 percent) to men . It turns out that only three out of 11 male employees receive zero from the male employers. It appears that male employers exclusively penalize the piece-rate choices of females employees. When men remunerate tournament choices, there is a sharp increase in the giving rate to women (35%).

5.4 Employee decisions

The employees make two decisions in *Choice* and *Feedback*: the remuneration choice and their effort decision which affects *performance*. In *Random*, they decide on effort and *performance* only. Note that *Choice* and *Feedback* are identical from the employees' perspective, so we do not distinguish between the two in this section.

Table 7 presents the employee's performance, that is, the average number of correctly solved problems. We focus on two main findings, both of which have been given attention in the previous literature. First, we find no significant gender differences in performance for both remuneration schemes (t tests, all possible combinations in the treatments with and without choice: $p = 0.145$ or higher). This is in line with Niederle and Vesterlund (2007). Second, we find that tournament employees significantly increase their performance under tournament conditions compared to piece rate. This difference is significant, taking all treatments into account (t -test, $p < 0.0001$) and also when considering the treatments with and without choice separately (t -tests, all $p < 0.021$).

	<i>Choice and Feedback</i>		<i>Random</i>	
	piece rate	tournament	piece rate	tournament
female	31.16 (10.40)	35.24 (12.30)	30.04 (8.72)	33.67 (9.30)
male	27.97 (9.92)	37.42 (11.31)	32.00 (10.03)	38.42 (12.65)

Table 7: Average number of correctly solved problems. Standard deviation in parentheses.

Table 8 presents an OLS regression analysis of employee performance. In regression (1) we include dummies for the treatments, piece rate, and the gender of the employee as independent variables. Regression (2) adds all interactions of *female employee* and the treatments and the compensation scheme. Except for *piece rate*, we observe no significant regressor. In particular, we do not find any significant performance differences between treatments or any gender effects. The negative sign of *piece rate* indicates that the performance of the employees is significantly higher under the tournament compared to piece-rate compensation, irrespective of the employee's gender or the treatment. We summarize:

Result 3. (i) No significant gender differences can be found regarding employee performance under piece-rate and tournament remuneration in our treatments. (ii) Employees perform significantly better under tournament than under piece-rate conditions in our treatments.

	performance	
	(1)	(2)
<i>choice</i>	-2.215 (1.600)	-1.159 (2.331)
<i>random</i>	-0.893 (1.591)	1.729 (2.315)
<i>piece rate</i>	-5.710*** (1.316)	-8.105*** (1.941)
<i>female employee</i>	-1.059 (1.307)	-0.0616 (3.098)
<i>female employee × choice</i>		-2.961 (4.257)
<i>female employee × random</i>		-5.533 (4.132)
<i>female employee × piece</i>		2.456 (3.846)
<i>female employee × random × piece</i>		2.026 (4.577)
<i>female employee × choice × piece</i>		2.781 (4.601)
<i>constant</i>	37.81*** (1.380)	37.53*** (1.656)
<i>R²</i>	0.082	0.097
<i>observations</i>	279	279

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8: OLS regression on subjects' performance.

We conclude this section with a discussion of the remuneration decisions. Around half of the employees in these two treatments chose tournament (52.2 percent). Men did so more often than women: 4 percent of male participants opted for tournament as opposed to 40 percent of female participants. This difference is statistically significant (2×2 Fisher’s exact test, $p = 0.002$).

Although we observe similar gender differences, the effect is somewhat less pronounced than in the previous literature. Niederle and Vesterlund (2007) find that 35 percent of their female participants but 73 percent of the males choose tournament. Several major differences of our experimental design may account for this. In addition to the fact that participants had no experience with the payment schemes in our case, it could be that some men are disinclined to compete in our setting where employers have full discretion over the final allocation of the money. Our findings can be summarized as follows:

Result 4. *Female employees choose tournament less often than men.*

6 Discussion

Our insights from previous studies (section 3) suggest the following hypothesis. Employers will reward tournament choices more generously because they can expect employees to perform better in the tournaments. The reward will, however, materialize in *Choice* only because employers have direct evidence on relative performance in *Feedback* and, in *Random*, subjects cannot be held accountable for the remuneration choices. We further expect employer decisions to be affected by a gender stereotype which should be particularly forceful in the absence of performance information.

Supporting this hypothesis, we find that the payment scheme affects employer behavior in *Choice* but not in *Random* and *Feedback*. Given a lack of information on relative performance, it appears that employers see tournament choices as a sign of good performance. For piece rate choices, they are affected by a gender stereotype. Stereotypical thinking may induce employers to believe that male employees perform better in the math task: given tournament was not chosen, the feedback “male employee” may be perceived as good news by the employers and so men accordingly receive higher rewards. In *Feedback* this effect indeed disappears, so providing employers with information on performance eliminates the misjudgment of female performance. The effect also disappears in *Random* where no remuneration choice is made.

To get a better idea of subjects’ (stereotype-affected) beliefs about employers’ performance, we ran a “bystander” treatment. An additional 45 subjects who did not take part in the main experiments were asked to read the instructions of the *Choice* treatment and to report their expectations of employees’ performances under *tournament* and *piece rate*. Each subject was asked three times

to guess the average performance of either female or male workers for the following three cases: piece rate, tournament winners, and tournament losers. The guesses were incentivized using a quadratic scoring rule.²² The data confirm the gender stereotype: male and female bystanders expect lower performances for female workers throughout.

While this gender bias regarding performance is intriguing, it cannot fully explain our results since the bias is present in all cases (piece rate, tournament winners, tournament losers). Our data, however, exhibit such a bias only in the case of piece-rate remuneration in *Choice* but not for *Random*.²³

A different explanation is based on employees' expectations of the money they would receive. Women who choose the piece-rate payment scheme may be seen as *expecting* a lower giving rate. Put differently, employers might think that women choosing tournament will demand a higher share of the pie; as do men for either payment scheme (Rigdon, 2012). To elicit employees' expectations, we asked a further 23 bystanders - who, as above, did not participate in the actual experiment - to read the instructions of the *Choice* treatment and report to us what they thought the employees' expectations were. Specifically, we asked these subjects whether, given a choice of either piece rate or tournament, women or men expect more (or whether they expect just as much).²⁴ People were paid an 8-euro flatrate for participation. Given a choice for piece rate, we find indeed that 11 of the 23 subjects suggest that men expect a higher giving rate; only four subjects assumed that women would be more demanding, and the rest thought both genders would be equally demanding. If employers are willing to meet these anticipated expectations, this would support the pattern we observe in the data; specifically, men obtain a higher share for piece rate in *Choice*. Having said that, the expectations data also suggest that men expect more when tournament is chosen (again, 11 subjects thought that men would expect a larger share of the pie). So men should also receive more when tournament is chosen - which is not the case.

We conclude as follows. The gender stereotype is consistent with our results only with the limitation that employers hold a mistaken belief about employee performance and misperceive the performance of men who choose piece rate. We find partial indication of that in our first bystander treatment. Bystanders also (wrongly) believe that female workers choosing tournament

²²We elicited these expectations after running our main treatments, so we did not ask the actual dictators themselves. However, using neutral bystanders for this task may even be advantageous since the actual dictators might exaggerate their estimates to justify their allocation decisions, perhaps mistakenly thinking the estimates would be communicated to the employees.

²³As an aside, we note that even if the "biased expectations" explanation had bite it would be entirely at odds with actual performances. The gender stereotype maintains that men *perform* better at the mathematical task, so ultimately it is the (allegedly) better performance that is rewarded, not male gender per se. In our data, men who choose piece-rate remuneration actually perform significantly worse than men under the tournament condition. And women outperform men (insignificantly) when working under the piece-rate condition. Hence, the feedback "male employee chooses piece rate" should be perceived as bad news by employers, not as good news.

²⁴To avoid a demand effect, we also asked whether they think people younger/older than 21 would expect more.

perform worse but, in the main experiment, women who choose tournament are not discriminated against. From our second bystander treatment, it also appears that giving rates are not exclusively expectation-based, either. Multiple and possibly contradictory forces may be at work here and further research is needed to come to a fully comprehensive understanding of the discriminatory effect we observe.

7 Conclusion

Our research is motivated by a simple but relevant question: if women and men differ in their inclination to compete, how do employers reward performances given the choice for either competitive or non-competitive remuneration? In our experiment, employees choose a payment scheme as in Niederle and Vesterlund (2007) but then employers, knowing the scheme chosen and the employee's gender, decide with full discretion on how much to give to the employee (like in a dictator game).

Our findings suggest that gender differences in the attitude to compete (Niederle and Vesterlund, 2007; Gneezy and Rustichini, 2004; Gneezy et al., 2009; Dohmen and Falk, 2011; Niederle et al., 2012; Buser et al., 2014) may cause a discriminatory gender wage gap. Women who decide to work under the non-competitive piece-rate remuneration scheme receive significantly less than men who make the same choice. By contrast, female workers who choose the competitive tournament are not paid less. Gender differences in payments also vanish when the payment schemes are randomly determined by the computer or when the employers are provided with the information on employees' relative performance. Women and men receive in all conditions a similar share of the pie, except when female employees chose the piece rate in our main treatment. Thus, it seems less women are rewarded for entering the tournament. Instead, women appear to be punished for choosing the non-competitive piece rate.

The wage gap we observe is discriminatory since it is adjusted for performance and payment scheme. In most countries such discrimination is illegal and there is equal pay for equal work by law. However, the adjusted wage gap in the field is, as a matter of fact, not zero in many countries (OECD, 2012). The implications of our experiments for the field have to be taken with a grain of salt: remuneration is often not a choice and information on relative performance may be available; also, our experiments are one-shot whereas decision makers in the field may gain experience. Nevertheless, our experiment suggests one possible channel for why adjusted wage gaps may occur. Whenever employers or superiors in general have discretion over wages, bonuses, promotions, perks, etc., the discrimination of women we observe may take effect.

References

- [1] Babcock, L., and Laschever, S. (2003) "Women don't ask: Negotiation and the gender divide." Princeton University Press.
- [2] Blau, F., D., and Kahn, L.M. (1992). "The Gender Earnings Gap: Learning from International Comparisons." *American Economic Review* 82, 533-538.
- [3] Blau, F., D., and Kahn, L.M. (1994). "Rising Wage Inequality and the U.S. Gender Gap." *American Economic Review* 84, 23-28.
- [4] Bohnet, I., van Geen, A., and Bazerman, M. (2012). "When Performance Trumps Gender Bias. Joint Versus Separate Evaluation." Working paper.
- [5] Bosman, R. and van Winden, F.A.A.M. (2002). "Emotional Hazard in a Power-to-Take Experiment." *The Economic Journal* 112 , 147-69.
- [6] Bowles, H., R., and McGinn, K., L. (2008). "Gender in Job Negotiations: A Two-Level Game." *Negotiations Journal* 24, 393-410.
- [7] Brock, J.M., Lange, A., and Ozbay, E.Y. (2013). "Dictating the Risk - Experimental Evidence on Giving in Risky Environments." *American Economic Review* 103, 415 - 37.
- [8] Buser, T., Niederle, M., and Oosterbeek, H. (2014). "Gender, Competitiveness, and Career Choices", *Quarterly Journal of Economics* 129, 1409-1447.
- [9] Cappelen, A.W., Hole, A.D., Sørensen, E.Ø., and Tungodden, B. (2007) "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review*, 97, 818-827.
- [10] Cappelen, A.W., Hole, A.D., Sørensen, E.Ø., and Tungodden, B. (2013) "Just Luck: An Experimental Study of Risk Taking and Fairness." *American Economic Review*, 103, 1398 - 1413.
- [11] Cherry, T.L., Frykblom, P., and Shogren, J.F. (2002). "Hardnose the Dictator." *American Economic Review* 92, 1218-1221.
- [12] Croson, R., and Gneezy, U. (2009). "Gender Differences in Preferences." *Journal of Economic Literature* 47, 448-474.
- [13] Dohmen, T., and Falk, A. (2011). "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101, 556-590.

- [14] Eccles, J.S., Jacobs, J.E., and Harold, R.E. (1990). "Gender role stereotypes, expectancy effects, and parents' socialization of gender differences." *Journal of Social Issues* 46, 183-201.
- [15] Eckel, C.C., and Grossman, P. J. (1998). "Are woman less selfish than men?: Evidence from dictator experiments." *The Economic Journal*, 108, 726-735.
- [16] Eckel, C.C., and Grossman, P. J. (2001). "Chivalry and solidarity in ultimatum games." *Economic Inquiry*, 39, 171-188.
- [17] Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). "Relative Earnings and Giving in a Real-Effort Experiment." *American Economic Review*, 101, 3330 - 48.
- [18] Fennema, E. H., Sherman, J.A. (1977). "Sex-related differences in mathematics achievements, spatial visualization, and sociocultural factors." *American Educational Research Journal* 14, 51-71.
- [19] Fennema, E. H., Sherman, J.A. (1978). "Sex-related differences in mathematics achievements and related factors: A further study." *Journal for Research in Mathematics Education* 9, 189-203.
- [20] Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual." *Experimental Economics* 10, 171-178.
- [21] Flory, A.F., Leibbrandt, and List, J.A. (2014). "Do competitive workplaces deter female workers? A large-scale natural field experiment on job-entry decisions." Forthcoming in *Review of Economic Studies*.
- [22] Fryer, R.G., and Levitt, D. (2010). "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics* 2, 210-240.
- [23] Gneezy, U., Leonard, K., List, J.A. (2009). "Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society." *Econometrica* 77, 1637-1664.
- [24] Gneezy, U., Niederle, M., and Rustichini, A. (2003). "Performance in Competitive Environments: Gender Differences", *Quarterly Journal of Economics* 118, 1049-1074.
- [25] Gneezy, U., and Rustichini (2004). "Gender and Competition at a Young Age", *American Economic Review Papers and Proceedings* 94, 377-381.
- [26] Greiner, B., (2004). "An online recruitment system for economic experiments." In: Kremer, K., Macho, V. (Eds.), *Forschung und wissenschaftliches Rechnen. 2003, GWDG Bericht 63. Ges. für Wiss. Datenverarbeitung, Göttingen*, 79-93.

- [27] Heinz, M., Juraneck, S., and Rau, H.A. (2012). "Do Women Behave More Reciprocally Than Men? Gender Differences in Real Effort Dictator Games." *Journal of Economic Behavior & Organization* 83, 105-110.
- [28] Hoffman, M., Gneezy, U., and List, J.A. (2011). "Nurture affects gender differences in spatial abilities." *PNAS*, published online August 29, 2011.
- [29] Hoffman, E., McCabe, and Smith, V. (1996). "Social Distance and Other-Regarding Behavior in Dictator Games", *American Economic Review* 86, 653-660.
- [30] Konow, J. (2000). "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90, 1072-1091.
- [31] Niederle, M., and Vesterlund, L. (2007). "Do Women Shy Away from Competition? Do Men Compete too Much?", *Quarterly Journal of Economics* 122, 1067-1101.
- [32] Niederle, M., Segal, C., and Vesterlund, L. (2012). "How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness." *Management Science*, 59, 1 - 16.
- [33] OECD (2012). "OECD Employment Outlook 2012".
- [34] Oxoby, R., J., and Spraggon, J. (2008). "Mine and yours: Property rights in dictator games." *Journal of Economic Behavior & Organization*, 65, 703 - 713.
- [35] Rigdon, M. (2012). "An Experimental Investigation of Gender Differences in Wage Negotiations", Working Paper.
- [36] Ruffle, B.J. (1998). "More is Better, But Fair is Fair: Tipping in Dictator and Ultimatum Games", *Games and Economic Behavior* 23, 247-265.
- [37] Spencer, S.J., Steele, C.M., and Quinn, D.M. (1999). "Stereotype Threat and Women's Math Performance", *Journal of Experimental Social Psychology* 35, 4-28.
- [38] Sutter, M., and Rützler, D. (2010). "Gender Differences in Competition Emerge Early in Life", *IZA Discussion Paper* 5015.
- [39] Swim, J.K. (1994). "Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes." *Journal of Personality and Social Psychology* 66, 21-36.
- [40] Swim, J.K. (1994). "Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes." *Journal of Personality and Social Psychology* 66, 21-36.

Appendix A

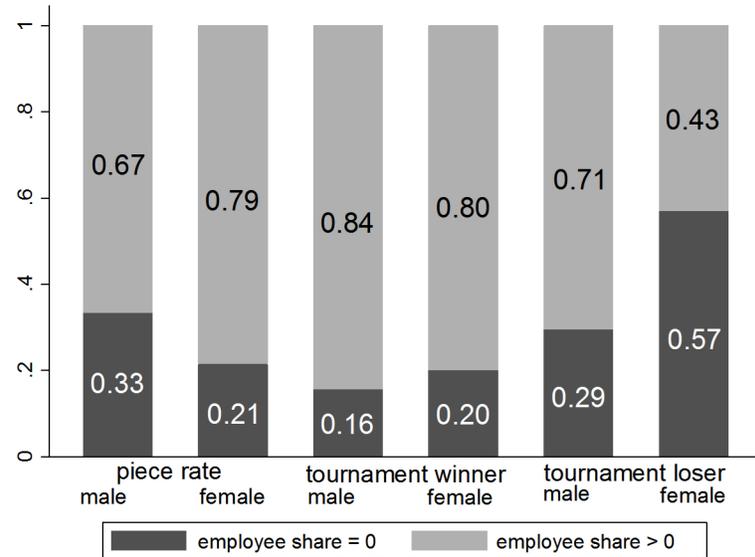


Figure 4: Distributions of giving rates conditional on remuneration choices and gender in *Feedback*.

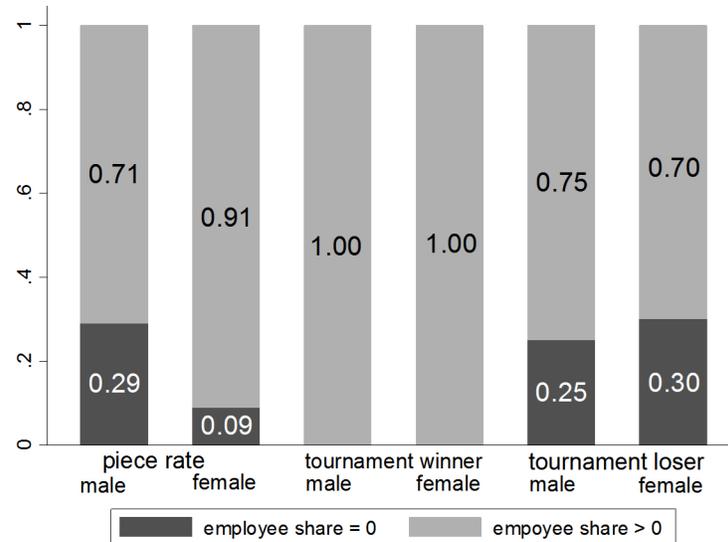


Figure 5: Distributions of giving rates conditional on remuneration choices and gender in *Random*.

Appendix B

Translated instructions in English (Treatment: *Choice*)

–not intended for publication–

–note: differences in the *Random* Treatment are presented in square brackets–

Information about the Experiment

General Information:

Please read these instructions carefully. For taking part in the experiment, you will get a single flat-rate amount of 5€, which is not dependent on your decisions in the experiment. During the experiment you can earn an additional amount in “Taler”; how much depends on your decisions and those of the other participants. This sum will be converted into Euros at the end of the experiment, added to the 5€, and paid out anonymously, always provided you completed all parts of the experiment. The Taler-to-euro conversion rate is 1.5 Taler = 1 cent.

The experiment is also anonymous, which means you will not learn who you are interacting with. Please turn off your mobile phone and do note that you may not speak to other participants during the complete experiment. If you have any questions, please raise your hand and we will come to you.

At the beginning of the experiment, you received a participant ID. It consists of a letter and a number. All participants with an “A” are type A participants. All participants with a “B” are type B participants. The number is your participant number, which is required at the end in order to pay you.

Note that the participants of both types, A and B, have been given the same instruction sheet.

Experimental Procedure:

In the context of the experiment each participant of type A is assigned to a participant of type B. Each participant B must split a sum of money between him/herself and the assigned participant A. The amount of money that participant B can allocate depends on the results that participant A achieved in a test run previously.

Instructions for Participants of type A

In a moment, the participants of type B will receive an envelope. In this they will find a piece of paper with an amount of money written on it, together with further information (more about this below). The amount of money depends on the result that you achieved in the preceding test (more about this below). Your allocated participant of type B must divide this between you and him/her. However, you get the lump-sum payment of 5€, regardless of the decision of participant B. Please keep in mind that the sum of money that was assigned to you by participant B will be paid together with the lump-sum payment at the end of the experiment.

The test: How does it work?

- The test is performed on the computer and is to calculate the sum of five randomly selected two-digit numbers.
- In total, you have 15 minutes for the test.

- You may not use a calculator.
- You will get scratch paper and pencils, which you are allowed to use to do the calculations.
- Using the computer keyboard, type your answer directly into the box “Sum” and click the button “Enter” with the mouse to submit it.

Please refer to the following Screen Shot that shows the structure of the input screen, as well as an example of a possible task. The computer will tell you at once if your answer was correct. Here too, your answers are kept confidential.

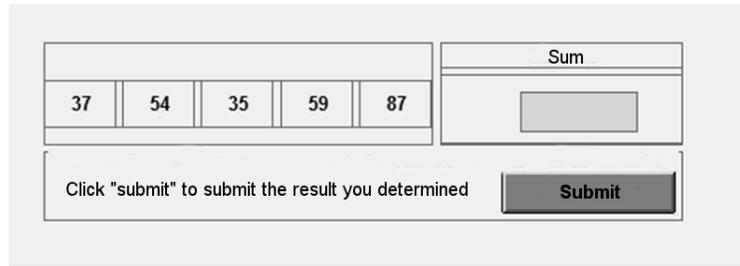


Figure 6: zTree Screen Shot of the math task (translated into English).

The test: Determining the sum of money

- During the experiment, the participants of type A are split into groups of six.
- The 6 groups have exactly 3 females and 3 males.
- The test lasts a total of 15 minutes.
- The amount of money is determined in two ways called piece rate and tournament.

Before the test begins you chose the payment type you would like. To do this, click on the initial computer screen on either “Piece rate” or “Tournament”. Then press “OK” and the experiment will begin. The selected method of payment applies for the entire test. The payment types are as follows:

[Random treatment: before the test begins the payment type is randomly set by the computer. This payment type applies for the entire test.]

Payment type: Piece rate

The sum of money achieved is based on the number of correctly answered items as calculated at the end of the 15 minutes. Each correctly answered task earns 50 Talers. Note: Only the correct answers count. Wrong answers do not reduce the sum of money. Thus for Piece rate:

Piece rate: The remuneration is as follows

earned money = number of correctly solved questions x 50 Talers

Payment type: Tournament

The sum of money achieved is based on the number of correctly answered items as calculated at the end of the 15 minutes. In contrast to payment type “Piece-rate,” the payment here depends on the relative performance compared to the other five participants of the group. There are two possible cases:

- Case 1: Occurs when you belong to the two participants in your group that have correctly answered the most tasks. In this case you are a winner of the tournament and get 90 Talers for each correctly answered task.
- Case 2: Occurs when you don't belong to the two participants in your group that have correctly answered the most tasks. In this case, you are not a winner of the tournament and you will get 30 Talers for every correctly answered task.

Thus, for Tournament:

Tournament: If you are a winner of the tournament (**case 1**), the remuneration is as follows

Earned money: number of correctly solved questions x 90 Talers

Tournament: If you are not a winner of the tournament (**case 2**), the remuneration is as follows

Earned money: number of correctly solved questions x 30 Talers

Note: Only the correct answers count. Wrong answers do not reduce the sum of money. If several participants correctly answered the same number of tasks, a random draw will be used, if needed. After completion of the test, we will distribute two questionnaires, which you will please fill out.

Examples:

- “Piece rate” selected by you [Random treatment: the computer set “Piece-rate” for you] and you answered 30 questions correctly.
The amount of money generated is then:
Amount = 30 x 50 Taler = 1500 Taler
In Euro = 1500 Taler / 150 = 10 Euro
- “Tournament” selected by you [Random treatment: the computer set “Tournament” for you], you answer 40 questions correctly and are therefore one of the two best participants of your group. You are paid according to case 1.
The generated amount of money is then:
Amount = 40 x 90 Taler = 3600 Taler
In Euro = 3600 Taler / 150 = 24 Euro
- “Tournament” selected by you [Random treatment: the computer set “Tournament” for you], you answer 30 questions correctly and you therefore do not belong to the two best participants of your group. You are paid according to case 2. The generated amount of money is then:
Amount = 30 x 30 Taler = 900 Taler
In Euro = 900 / 150 = 6 Euro

Instructions for participants of type B

For explanations of this test, please refer to the sections “The test: how does it work” and “test: determining the sum of money” on pages 2 + 3 of these instructions.

- Please ensure that your assigned participant of type A has completed all the necessary fields. Then enter your participant-ID in the field provided.
- Decide on the allocation of the total amount between you and your assigned participant of type A:
Enter your chosen distribution on the sheet of paper provided for this.

Please keep in mind that you can choose all splits between 0 Talers and the total amount of money achieved in Talers (please enter only whole numbers, no decimal points). The following are a few examples of possible splits:

1. Example: Amount of money: 3,000 Talers
I keep 3,000 Talers.
I leave participant A 0 Talers
2. Example: Amount of money: 1,500 Talers
I keep 750 Talers.
I leave participant A 750 Talers
3. Example: Amount of money: 1,500 Talers
I keep 100 Talers.
I leave participant A 1,400 Talers

Please note that these amounts will be paid at the end of the experiment together with the lump-sum payment (5€).

Translated instructions in English (additional treatment 1: *Employees' performance*)

–not intended for publication–

Note: In the *performance* treatment we first handed out the full instructions of *Choice*. We then distributed three additional questions where we asked the participants about employees' performance. Below, we list these three questions. Note, there were two versions of treatments. In each version we either ask all three questions for male or for female employees. That is, why we present the questions in the form: "female/male".

[.]

Below are examples of possible remuneration schemes employees could have opted for. Your task is now to assess average performance of the employees under the different conditions.

What do you think:

1. *What is the average number of correctly solved questions by female/male participants who opted for piece rate ?*
2. *What is the average number of correctly solved questions by female/male participants who opted for tournament and won it ?*
3. *What is the average number of correctly solved questions by female/male participants who opted for tournament and lost it ?*

