The boys crisis: Do we consider it more acceptable when males fall behind than when females do?

Preliminary version.

Alexander W. Cappelen, Ranveig Falch and Bertil Tungodden

July 11, 2016

Abstract It is well established that there is a gender bias in top-level jobs where males dominate. However, there is also an increasing worry of a gender bias in the lower tale of education and labor market outcomes, often termed the 'boys crisis'. What can explain these patterns? In this project, we study experimentally whether people consider it more acceptable when males fall behind than when females do, using a novel design implemented on a representative sample of Americans. The participants make distributive choices involving males and females and we randomly manipulate the gender composition in the distributive situations. We show that people find it more acceptable when males fall behind than when females do when outcomes reflect merit. We provide evidence showing that this result is not driven by a general preference for females, but is specific to how people redistribute in merit environments where males perform worse than females. We argue that this finding may shed light on the gender discrimination against males in different parts of society, on the 'boys crisis' and why males increasingly lag behind females in merit settings such as in education and in the lower tale of the labor market.

Key words: inequality, fairness, responsibility, gender discrimination, experiment

JEL code: C91, D63, J16

The study was organized by The Choice Lab. We would like to thank Caroline Bonn and Matthias Sutter for excellent comments. We are also thankful the financial support from the NORFACE (New Opportunities for Research Funding Agency Cooperation in Europe) Network and The Centre for Ethics and Economics at the Norwegian School of Economics.

1 Introduction

Across the world, we observe that males occupy the majority of top level jobs and heads of government positions (Bertrand and Duflo, 2016; Bertrand and Hallock, 2001). At the same time there is an increasing worry of a gender bias in the lower tale of outcomes, often termed the 'boys crisis'. The 'boys crisis' is characterised by males lagging behind females in high school and college attainment in many developed countries and males with less than four-year college education in the US having had a reduction in real income over the last decades (Autor and Wasserman, 2013). What can explain this pattern?

Accumulated evidence suggests that in developed countries, teacher grade setting is discriminatory against males in a range of subjects in kindergarten, primary school, high school and higher education (Breda and Ly, 2015; Cornwell, Mustard, and Van Parys, 2013; Falch and Naper, 2013; Lavy, 2008; Lindahl, 2015; Robinson and Lubienski, 2011; Terrier, 2015). This gender-bias against males in the education system may be part of the reason for the 'boys crisis' and while there may be several other explanations for the 'boys crisis' we want to investigate this one further, and see whether this gender-bias is particular to the educational setting, or if it is reflected in mixed-sex environments in general. More specifically, we ask whether people consider it more acceptable when males fall behind than when females do and study this experimentally in an incentivised, large-scale controlled lab setting. To our knowledge, we provide the first study addressing this question and we offer a novel experimental design implemented on a representative sample of Americans. The participants make distributive choices in mixed-sex environments, where we randomly manipulate whether the male or the female has more earnings. We also study redistributive behavior in single-sex environments, allowing us to study whether people to a greater extent hold stakeholders responsible for outcomes in single-sex distributive situations only involving men, compared to single-sex distributive situations only involving women. The previous literature on fairness preferences (Almås, Cappelen, and Tungodden, 2015; Bolton and Ockenfels, 2000; Cappelen, Drange Hole, Sørensen, and Tungodden, 2007; Charness and Rabin, 2002; Engelmann and Strobel, 2004; Fehr, Bernhard, and Rockenbach, 2008; Konow, 2000; Sutter, 2007) has demonstrated that the source of inequality is of great importance for people's willingness to redistribute. Informed by these findings our design includes both distributive situations where luck is the source of inequality and distributive situations where merit is the source of inequality.

Our study adds to the growing literature on how fairness preferences shape distributive behavior (Almås, Cappelen, Sørensen, and Tungodden, 2010; Bolton and Ockenfels, 2000; Cappelen et al., 2007; Fehr and Schmidt, 1999; Konow, 2000; Konow, Saijo, and Akai, 2009) by studying whether people's fairness ideals differ across contexts.[1] In contrast to the previous studies, we are interested in whether we use different fairness ideals in different environments. We use a spectator design, since spectator decisions provide a direct expression of the moral preferences of the participants (Konow et al., 2009; Cappelen, Moene, Sørensen, and Tungodden, 2013). Finally, our results also shed light on the large and diverse literature on gender discrimination (see e.g. Bertrand and Duflo (2016); Bertrand (2011); Black and Strahan (2001); Castillo, Petrie, Torero, and Vesterlund (2013); Dittrich et al. (2015); Goldin and Rouse (2000); Niederle, Segal, and Vesterlund (2013); Reuben, Sapienza, and Zingales (2014); Sharma (2015)). Bertrand and Duflo (2016) reviews the field experiment literature on discrimination and describes common tools psychologists use to study discrimination in the lab. They find that while there are many field experiments on discrimination, in particular audit and correspondence studies, there has been a struggle within economics to identify the driving mechanisms of discrimination and in particular linking the data to specific economic theories on discrimination, such as taste-based and statistical discrimination. One attempt is made by List

[1] Previous studies have varied the salience of recipient gender in stakeholder games such as the dictator game and the ultimatum game. For an overview, see e.g. Dittrich, Büchner, and Kulesz (2015); Sharma (2015).

(2004) who matches data from a field experiment on racial discrimination in a sports cards market with the results from lab experiment where the card dealers play a dictator game to rule out taste-based explanations. However, although he does not find evidence of discrimination in the dictator game, this alone cannot prove that there is no taste-based discrimination in the field part of the experiment. Bertrand and Duflo (2016) further describes how psychologists have studied discrimination in the lab, often using implicit association tests, Goldberg Paradigm experiments and list randomizations, and how psychologists have developed their own strain of literature on the micro-foundations of discrimination, including "personality development, socialization, social cognition, evolutionary psychology and neuroscience" (p. 4 - om dette er interessant aa ta videre kan side 4-5 i artikkelen vaere fint aa lese gjennom), somehow blurring the sharp line between taste-based and statistical discrimination. In addition, the vast majority of these field experiments does not focus on gender.

Our main findings shows that males are indeed discriminated against i mixed-sex merit environments in the same way as in the educational setting. Figure 1 shows the standardized transfers to the losing worker in two mixed-sex environments where merit is the source of inequality. When a female loses to a male the spectators redistribute 0.174 standard deviations more to her than what they do to the male losing to a female.

[Figure 1 about here]

Figure 2 splits the spectators in two groups, female and male spectators, before showing the standardized transfers to the losing worker in two mixed-sex environments where merit is the source of inequality. We find that male spectators are not gender-biased, while the female spectators on average redistribute 0.284 standard deviations more to the female loser compared to the male loser.

[Figure 2 about here]

Structure?

2 Sample

To collect experimental data on a nationally representative sample, we combine the infrastructure of an international labor market and the infrastructure of a leading international data-collection agency to run a real effort dictator game with a spectator design (Almås et al., 2015).

There are two types of participants in the experiment, *workers* and *spectators*. We carried out two rounds of data collection. In the first round, we recruited 1370 workers from the international online labor market Amazon Mechanical Turk (mTurk), 685 men and 685 women living in the US. In the second round, we recruited 702 workers; 351 men and 351 women. The spectators in the experiment were recruited using the data-collection agency TNS Gallup. In the first round of data collection, we recruited 2052 US participants who are nationally representative (18+ years old) on observable characteristics (gender, age and geography). In the second round, we recruited another 1050 nationally representative US participants. Individuals who participated in the first round of data collection were not be permitted to participate in the second round. Table 1 provides an overview of the background characteristics of the spectators.[2]

[Table 1 about here]

[2]The median yearly household income in the US was $53,657 in the 2014 US census (DeNavas-Walt and Proctor, 2015) and the median age was abut 37.2 in the 2010 US census, having increased by two years during since 2000 (Howden and Meyer, 2011).

3 Experimental design

We conducted a version of a real effort dictator game with a *spectator* design. The spectators made a choice that had monetary consequences, not for themselves, but for two *workers* who had completed a real effort assignment. In the following, we explain in more detail the design and the instructions given to the workers and the spectators.

3.1 Workers

The workers signed up for the experiment at the mTurk website. They completed three real effort assignments, but made no distributive choices. For each assignment, each worker was matched with another worker who had also completed the assignment, and the two constituted a pair that in turn was matched with a spectator. The workers received a fixed payment of 2 USD for taking the HIT as well as a bonus payment averaging 3 USD per assignment based on either luck or merit and on the spectator choice explained in the following section. Importantly, except for the fixed participation fee, the workers were only informed how their earnings would be determined after they had completed the assignments. Since the behavior of the workers is not essential for the present study and since the spectators were not informed about the type of tasks assigned to the workers or their productivity distribution, we only provide a discussion of the workers' instructions in relation to the choices made by the spectators.[3]

3.2 Spectators

The spectators made an incentivized distributive choice and answered a set of standard background questions. We now explain each part in detail.

3.2.1 Distributive choice

The spectator decided on the payment for a pair of workers. Each spectator only considered one distributive situation, meaning that we had a between-individual design, and all spectators faced the same pre-redistribution earnings of 0 USD, 6 USD. They also knew that the workers did not receive information about how their earnings would be determined until after completing the assignments, and that the workers would not receive any feedback except for their payments. Table 2 summarizes the main stages in the experiment.

[Table 2 about here]

In the first round of data collection we further implemented a 2x2x2 design where we varied whether;

- whether the pair of workers is mixed-sex or single-sex,
- whether the source of inequality is luck or merit, and

[3]However, in case of interest, Figure A.1 and A.2 present the gender distributions of the worker productivity in the two types of worker tasks; a counting task and a code recognition task. We also test whether male and female workers performed differently on each of the two tasks by performing a two-sided t-test checking whether the means in the two groups were different and by using a Kolmogorov-Smirnov test (with permutation) to check whether the distribution of results in the two groups were different. In the counting task, there were no significant differences between male and female performance (t-test: mean error males 162.41, mean error females 154.53, p-value=0.2795. Kolmogorov-Smirnov test: No significant difference between the male and female distribution (p-value=0.2759)). However, in the code recognition task, females performed slightly worse than males on average (t-test: mean score males 88.72, mean score females 86.59, p-value=0.0279) and their two distributions were significantly different from each other (Kolmogorov-Smirnov test with p-value=0.0375).

- whether the (pre-distribution) unlucky or least productive worker is a man or a woman.

This 2x2x2 design gave us eight treatments that allow us to study whether gender composition matters for redistributive decisions, both in the context of luck and merit. The eight treatments are presented in Table 3.[4] The exact instructions given to the spectators in treatments T1 and T8 are provided in A.1. The six other treatments are all variations over these two with respect to gender and the source of inequality.

[Table 3 about here]

In the second round of data collection, we only collected data for Treatments 3 and 4, i.e. the mixed-sex merit treatments. The instructions for these two treatments were exactly the same as in the first round of data collection.

3.2.2 Belief and Attitude questions

To study underlying mechanisms, we included two questions on beliefs and attitudes in the second round of data collection. These questions were asked after the spectators had made their distributive decisions.

Question 1) US 8th graders were tested in mathematics and reading. How do you think the male students performed relative to the female students in

- a) mathematics?
  - Males much better
  - Males somewhat better
  - Equal performance
  - Females somewhat better
  - Females much better

We then asked the same question about reading (1b), by inserting reading instead of mathematics in the question structure above.[5]

Question 2) Do you generally favor or oppose affirmative action programs for women?[6]

- Generally favor
- Generally oppose

[4]Table A.1 presents a balance test where we use joint F-tests to test whether the eight treatments are significantly different from each other with respect to any of the background characteristics in Table 1. We control for round of data collection and find no evidence of significant variation in the background characteristics across treatments.

[5]Question 1a and 1b are based on a nationally representative assessment conducted in 2015 by The National Assessment of Educational Progress (NAEP). Here, male and female 8th graders performed equally well on the mathematics test (each group with an average of 282 points), while females on average performed slightly better than males on the reading test (271 vs. 260 points). For more details, see http://www.nationsreportcard.gov/reading_math_2015/.

[6]Question 2 is taken from Gallup's Minority Rights and Relations survey conducted in 2015 with more than 2000 US adults. They found that 67% of Americans are in favor of affirmative action for women, with females being more prone to support it than males (72% vs. 62%).

3.2.3 Background questions

In both rounds of data collection, the spectators were asked several background questions. These included questions on age, gender, what region they live in, yearly household income, educational background and political orientation. A complete list of the background questions with exact formulations are provided in A.2. The questions were asked after the distributive decisions and the belief and attitude questions.

4 Research question

The experiment is designed to study whether there is gender discrimination in the assignment of responsibility for outcomes. Do we consider it more acceptable when males fall behind than when females do? We study this in mixed-sex environments and investigate it both in the context of luck and merit.[7] A large literature has focused on the gender discrimination of females in various context, but there is also evidence suggesting that males sometimes are discriminated against (Cornwell et al., 2013; Lavy, 2008; Lindahl, 2015). The present project provides a novel test of whether males or females are discriminated against in the assignment of responsibility for outcomes.

5 Results

This section provides the results. A discussion of the results is presented in section 6.

5.1 The distributive decisions

Figure 3 gives an overview of the spectator decisions, illustrating the mean dollar amount transferred to the losing worker across all treatments. About one third of the spectators do not redistribute anything, while another third implements equality, redistributing 3 USD to the losing workers. As we should expect, only a few spectators redistribute more to the worse off than to the better off workers, implying that the final third of the spectators redistribute 1 or 2 USD to the losing one.[8]

[Figure 3 about here]

5.2 Main findings

We find strong evidence of males being discriminated against in the assignment of responsibility when merit is the source of inequality. Table 4 below provides the results from the mixed-sex merit treatments, focusing both on our main outcome variable; the standardized amount transferred to the losing worker (columns 1 and 2), and on the following additional measure; the share of spectators giving nothing to the losing worker (columns 3 and 4).

The following regression is the basis for the results presented in Table 4:

$$u_i = \alpha + \beta_1 Maleloser_i + \gamma_1 X_i + \gamma_2 Round2_i + \epsilon_i, \tag{1}$$

where u_i is the standardized transfer to the losing worker (standardized with respect to the two mixed-sex merit treatments), $Maleloser_i$ is an indicator taking the value 1 if individual i belongs

[7]The project is also designed to study whether we hold males more responsible for outcomes than females in single-sex environments. Our pre-analysis plan for the first round of data collection put forward two hypotheses to be tested both in the context of luck and merit; one for the single-sex environments and one for the mixed-sex environments. Hypothesis 1: *Males are held more responsible for outcomes than females.*. Hypothesis 2: *There is no gender discrimination in the assignment of responsibility.*. The results for both hypotheses are presented in the paper, but the main focus will be on whether there is gender discrimination in the assignment of responsibility.

[8]To get an overview of the distributive decisions by treatment and round, see Figures A.3 and A.4.

to a treatment where a male lost, $Round2_i$ is an indicator taking the value one if individual i participated in the second round of data collection and ϵ_i is an error term. We include extra background variables in columns 2 and 4 in Table 4. These make out X_i and includes (as specified in our pre-analysis plan) indicator variables for gender, age, political orientation and income that are equal to one if spectator i is a male, is below the median age in the sample, is a republican and has an income below the median in the US, respectively. The sample includes the two mixed-sex merit treatments (treatments 3 and 4). In columns 3 and 4, the dependent variable, u_i, is switched out with an indicator variable that takes the value one if the spectator gave nothing to the losing worker.

[Table 4 about here]

We find that the average amount redistributed to the less productive worker is reduced by 0.174 standard deviations when the losing worker is a male rather than a female (p=0.001) (Table 4, column 1). This result remains practically unchanged when we include background variables for gender, age, political preference and income (Table 4, column 2). If we use the alternative outcome variable (columns 3 and 4) we find the same qualitative results: The share of spectators giving nothing to the losing worker increases by 7.3 percentage points when the losing worker is a male rather than a female (p=0.003) The result remains unchanged when we include the additional background variables.[9]

5.3 Mechanisms

In this section we investigate the mechanisms that can explain that the spectators consider it more acceptable when males fall behind than when females in contexts of merit. In order to do this, we provide the results from the final six treatments, a heterogeneity analysis including gender, age, political preference and income (as specified in both of our pre-analysis plans), an analysis of the belief and attitude questions and and analysis investigating how the different mechanisms interact.

5.3.1 The mixed-sex luck and single-sex treatments

We use regression 1 to analyse the six remaining treatments. Table 5 provides the results. The standardization of the main outcome variable in based on the respective two treatments under analysis (mixed-sex luck in columns 1 and 2, single-sex merit in columns 3 and 4 and single-sex luck in columns 5 and 6).

[Table 5 about here]

We find that there is no gender differentiation in the mixed-sex environment where luck is the source of inequality (Table 5, column 1, p=0.705). The result remains the same when we include the additional background variables in column 2. We do not find a difference in transfers to the losing worker in either of the single-sex environments either. In the single-sex merit setting (columns 3 and 4), there is no significant difference in how much the spectators on average transfer to the losing worker in a male single-sex setting and a female single-sex setting (column 3, p=0.647). The same is the case in the single-sex luck setting (columns 5 and 6) where there is no difference in how the spectators hold males and females responsible (column 5, p=0.325), and the results remain

[9] For robustness check, we provide the results from running regression 1 by rounds of data collection (Table A.2). The sample is still spectators in treatments 3 and 4.

practically the same if we include additional background variables.[10][11]

5.3.2 Heterogeneity analysis

In the heterogeneity analysis we ask whether the observed difference in distribution between the two mixed-sex merit treatments is particularly strong in some subgroups.[12] As specified in the pre-analysis plan, we perform the heterogeneity analysis using the following background variables and specifications: gender (male or female); age (below or above the median age in the representative US sample (18+ years)), political orientation (republican, non-republican); and income (below or above the median in the US). The regression results are provided in Table 6 in column 1, 2, 3 and 4, respectively.[13]

$$u_i = \alpha + \beta_1 Maleloser_i + \beta_2 M_j \times Maleloser_i + \gamma X_i + \epsilon_i, \tag{3}$$

where u_i is the standardized amount transferred to the losing worker (standardized with respect to the two mixed-sex merit treatments), $Maleloser_i$ is an indicator taking the value one if individual i belongs to a treatment where a man lost, $M_j \times Maleloser_i$ is an interaction term taking the value one if M_j is equal to one (which is the case when spectator j is a male) and if individual i belongs to a treatment where a man lost, X_i includes indicator variables for gender, age, political orientation and income that are equal to one if spectator i is a male, is below the median age in the sample, is a republican and has an income below the median in the US, respectively, and ϵ_i is an error term. We will use corresponding regression equations for the other dimensions of heterogeneity. We regress 3 on the sample of spectators in treatments three and four (the mixed-sex merit treatments).

[Table 6 about here]

Gender: β_1 is negative and significant, meaning that female spectators on average gender discriminate males in the mixed-sex merit environment (β_1 is equal to -0.284 standard deviations with a p<0.000). β_1+β_2 is on the other hand insignificant, implying that male spectators on average

[10]For robustness, we use the following regression to conduct a joint F-test. We test whether there is any significant treatment variation in transfers between the mixed-sex luck, and the single-sex treatments except for the increase in transfers we should expect going from a context of merit to luck.

$$u_i = \alpha + \beta_1 Maleloser_i + \beta_2 Lucksingle_j + \beta_3 Meritsingle_j + \beta_4 Lucksingle_j \times Maleloser_i + \beta_5 Meritsingle_j \times Maleloser_i + \gamma X_i + \epsilon_i, \tag{2}$$

where u_i, $Maleloser_i$, X_i and ϵ_i are the same as in regression 1, $Lucksingle_j$ is an indicator variable taking the value one if the spectator belongs to either of the two single-sex luck treatments (treatments 5 and 6), $Meritsingle_j$ is an indicator variable taking the value one if the spectator belongs to either of the two single-sex merit treatments (treatments 7 and 8), $Lucksingle_j \times Maleloser_i$ is an interaction term equal to one if both $Lucksingle_j$ and $Maleloser_i$ are equal to one and $Meritsingle_j \times Maleloser_i$ is an interaction term equal to one if both $Meritsingle_j$ and $Maleloser_i$ are equal to one. The sample is restricted to the mixed-sex luck and single-sex treatments (1,2 and 5-8). Table A.3 provides the results (column 1 without backround variables, column 2 with) of the regression and the results of the joint F-test testing whether the following variables are jointly significant from each other $Maleloser_i = Lucksingle_j \times Maleloser_i = Meritsingle_j \times Maleloser_i$. They are not (P-value=0.5923 in column 1 and 0.5623 in column 2), which means that the F-test supports our previous findings that the interesting treatment variation is to be found in the mixed-sex merit environments and not in the remaining six treatments.

[11]The lack of gender differentiation in the single-sex luck and merit environments (columns 3-6 in Table 5) also means that we find no evidence to support for Hypothesis 1 (specified in the pre-analysis plan for the first round of data collection): *Males are held more responsible for outcomes than females.* This is true in both of the single-sex environments, when the source of inequality is merit and when it is luck.

[12]We also perform a heterogeneity analysis for the remaining six treatments. These results are reported in Tables A.4, A.5 and A.6.

[13]In case of interest, you may find the same results in Table A.7 where all of the background variables are displayed (except for the variable indicating round of data collection).

do not gender discriminate in the same environment (p=0.434). β_2 shows that the two effects are significantly different from each other (p=0.024).[14]

Age: The median age in our sample is 41, and we split the sample of spectators above and below the median age to analyse the two groups separately. We find that both spectators above and below the median age do gender discriminate in the mixed-sex merit environments (β_1=-0.212 with a p=0.003 and $\beta_1+\beta_2$=-0.134 with a p=0.061), and β_2 shows that the two groups are not significantly different from each other (p=0.437).

Political orientation: We further split our sample between republicans and non-republicans. We find that both republican and non-republican spectators gender discriminate in the mixed-sex merit environments (β_1=-0.169 with a p=0.006 and $\beta_1+\beta_2$=-0.181 with a p=0.039), and β_2 shows that the two groups are not significantly different from each other (p=0.910).

Income: Lastly, we split our sample into groups according to income level, the spectators above and below the median income level in the US. We find that β_1 is negative and significant, meaning that spectators with above median income on average gender discriminate males in the mixed-sex merit environment (β_1 is equal to -0.227 with a p=0.001). $\beta_1+\beta_2$ is on the other hand insignificant, implying that spectators with a below median income on average do not gender discriminate in the same environment (p=0.230). However, β_2 shows that the two groups are not significantly different from each other (p=0.195).

Heterogeneity analysis, robustness: Additionally, we run 3 on all four interaction terms (each background term interacted with Male loser) to check the robustness of the results from the heterogeneity analysis. We find that female spectators are significantly more prone to gender discriminate against males even when controlling for all heterogeneity effects (Male*Male loser=0.229 with a p=0.023). These results are shown in Table 6, column 5.[15]

5.3.3 Underlying mechanisms: Beliefs and attitudes

In the second round of data collection we included a set of belief and attitude questions to study the underlying mechanisms of the gender discrimination of males in the mixed-sex merit environments. The gender discrimination that we observe is consistent with classical gender stereotyping where females are considered less able than males (Fiske, Cuddy, Glick, and Xu, 2002; Eagly and Karau, 2002; White and Gardner, 2009). In the mixed-sex merit environment, a possible mechanism for the gender discrimination we observe could therefore be that a less productive male is interpreted as reflecting very low effort or talent and correspondingly that a female ahead is interpreted as reflecting extraordinary effort or talent. From previous experimental literature it is established that the majority of participants hold people responsible for their effort and talent (Cappelen, Sørensen, and Tungodden, 2010; Cappelen et al., 2007; Konow, 2000), which means that such a mechanism would lead to males being held more responsible than females for lagging behind.

An alternative interpretation of the gender discrimination is that it may reflect affirmative action towards high performing women in mixed-sex environments. Affirmative action is the policy

[14] We also provide the results from regression 1 split by the gender of the spectators (still from treatments 3 and 4). These results are reported in Table A.9 and remain largely the same as in the recently discussed heterogeneity analysis for gender, where we use regression 3.

[15] We do the same analysis including an additional background variable that was not specified in the pre-analysis plans. This variable is low education, an indicator variable which is equal to one if the spectator stopped his/her education at age 19 or younger. We use 3 on the sample of spectators in treatments 3 and 4 again, and find no evidence of the treatment effect varying between the two education groups (see Table A.8. The results of the main heterogeneity analysis (Table 6) remain practically unchanged to the inclusion of the extra background variable.

of favoring members of a disadvantaged group, promoting their rights, opportunities or progress.[16] If women are considered a disadvantaged group relative to men, then a preference for affirmative action may cause participants to favor females ahead more than males ahead.

We investigate both of these proposed explanations based on the belief and attitude questions described in section 3.2.2.

Testing belief and attitude questions: To test whether people who overrate the performance of males relative to females, are more prone to gender discriminate against males in the assignment of responsibility for outcomes we create a dummy for those who overrate the performance of males relative to females in mathematics, defined as an indicator variable which is equal to one for spectators who overrate males relative to females on belief question 1a. To test people who generally favor affirmative action programs for women are more prone to gender discriminate against males in the assignment of responsibility for outcomes, we create an indicator variable that is equal to one for spectators who generally favor affirmative action programs for women, and zero for those who generally oppose.[17]

To check whether people who overrate the performance of males relative to females, are more prone to gender discriminate against males in the assignment of responsibility for outcomes, we regress the following equation on the data collected in the second round:

$$u_i = \alpha + \alpha_d Overrate(math)_j + \beta_1 Maleloser_i + \beta_{d1} Overrate(math)_j \times Maleloser_i + \epsilon_i, \quad (5)$$

where u_i is the standardized transfer to the losing worker (standardized with respect to the two mixed-sex merit treatments), $Overrate(math)_j$ is an indicator taking the value one if spectator j overrates the performance of males relative to females on Question 1a, $Maleloser_i$ is an indicator taking the value one if individual i belongs to a treatment where a male lost, $Overrate(math)_j \times Maleloser_i$ is an interaction term taking the value one if $Overrate(math)_j$ is equal to one and if individual i belongs to a treatment where a male lost, and ϵ_i is an error term.

[Table 7 about here]

Looking at column 5, Table 7, we find that the spectators who overrate males relative to females in mathematics discriminate against males in the mixed-sex environment ($\beta_1+\beta_{d1}$ is negative and significant, -0.206 (p=0.046)). However, we find the same for those who do not overrate males relative to females (β_1 is equal to -0.156 (p=0.043)), and the two groups are not significantly different from each other with respect to gender discrimination (β_{d1} is equal to -0.050 (p=0.696)).

[16] For a survey of the literature on the effects of affirmative action, see Holzer and Neumark (2000). For papers on affirmative action of women in the lab, see e.g. Balafoutas and Sutter (2012) and Niederle et al. (2013).

[17] Before both of these analyses we performed balance tests, checking whether there were treatment effects on belief question about math (Question 1a) or on the affirmative action question (Question 2). We tested this by regressing the following equation on the data collected in the second round:

$$Overrate(math)_j = \alpha + \beta_1 Femaleloser_i + \epsilon_i, \quad (4)$$

where $Overrate(math)_j$ is an indicator taking the value one if spectator j overrates the performance of males relative to females in mathematics (Question 1a), $Femaleloser_i$ is an indicator taking the value one if individual i belongs to the mixed-sex treatment where a woman was least productive, and ϵ_i is an error term. We find that the estimated value of β_1 is insignificant (p-value=0.247), which means that there is no treatment effect on the answer to the belief question concerning math (Question 1a). We provide the same test for the affirmative action question (Question 2), using 4 where we replace $Overrate(math)_j$ with the indicator $Aff.ac._j$, which takes the value one if spectator j favors affirmative action programs. We again find that the estimated value of β_1 is insignificant (p-value=0.839), which means that there is no treatment effect on the attitude question about affirmative action (Question 2). In both cases, we find that the belief and attitude questions are not affected by treatment, and we consider them as valid background variables to be used in the interaction analysis.

These results hold when controlling for the spectators' gender, age, political preferences and income (column 6).

We provide the same test for the attitude question concerning affirmative action (Question 2) to check whether people who generally favor affirmative action programs for women are more prone to gender discriminate against males in the assignment of responsibility for outcomes. Again, we use regression (5) and replace $Overrate(math)_j$ with the indicator $Aff.ac_j$, which takes the value one if spectator j stated that he/she generally favors affirmative action programs for women and $Overrate(math)_j \times Maleloser_i$ with $Aff.ac_j \times Maleloser_i$, which is equal to one if both $Aff.ac_j$ and $Maleloser_i$ are equal to one. The results from this analysis are also provided in Table 7. Looking at column 2, we find that the spectators who favor affirmative action programs for women discriminate males in the mixed-sex environment ($\beta_1 + \beta_{d1}$ is negative and significant, -0.260 ($p<0.000$)). Interestingly, we do not find this to be the case for the spectators who oppose affirmative action for women (β_1 is equal to 0.029 (p=0.797)), and we find that the two groups are significantly different from each other with respect to gender discrimination (β_{d1} is equal to -0.289 (p=0.032)). These results hold when controlling for the spectators' gender, age, political preferences and income (column 3).

Additionally, we regress (5) with the inclusion of $Overrate(math)_j$, $Aff.ac._j$, the two corresponding interaction terms with Male loser and the set of background variables to check the robustness of the results from the mechanisms analysis. The results are provided in Table 7, column 7 and we find that spectators who favor affirmative action for women are significantly more prone to gender discriminate against males even when controlling for the other underlying mechanism.[18]

5.3.4 Heterogeneity analysis with affirmative action

In this section we test whether including affirmative action alters the results of the heterogeneity analysis we presented in 5.3.2 and Table 6. We again use regression 3, but include the indicator variables $Aff.ac._j$ and $Aff.ac_j \times Maleloser_i$ together with the gender analysis (column 1) (the variables are defined in section 5.3.3). Table 8 provides the results.

[Table 8 about here]

Skal egentlig aff.ac. vaere med i lincomene her (tabell 8)? Jeg er noe usikker paa tolkningen av $Maleloser_i + Aff.ac_j \times Maleloser_i$ og $Aff.ac_j \times Maleloser_i$ under.

What we find is that when we include affirmative action, the gender effect on discrimination of males goes away (column 1) while the large heterogeneity in gender discrimination remains between

[18]We also split the sample of spectators from the second round of data collection into females (Table A.10) and males (Table A.11) to investigate whether the results above are driven by one specific group. Looking at column 5, we find that female spectators who overrate males relative to females in mathematics gender discriminate against males in the mixed-sex merit environment (p-value=0.034). However, this is also here true for those who do not overrate (p-value=0.019), and the two groups are not significantly different from each other (p-value=0.752). The same is true when we include background variables in column 6. Looking at column 2, we find that female spectators who generally favor affirmative action for women gender discriminate against males (p-value=0.001), and that this is not the case for those who oppose affirmative action (p-value=0.631). The two groups are significantly different from each other (p-value=0.202), and the result holds when controlling for background variables (column 3). Consequently, the results for the female spectators are qualitatively the same as the sample of both males and females. The same is not true for the males (Table A.11). Neither the males who overrate males relative to females in mathematics nor the males who do not overrate gender discriminate in the mixed-sex merit environment (p-values=0.426 and 0.557, respectively) and the two groups are not significantly different from each other (p-value=0.780) (column 5). The same is true if we control for background variables (column 6). Looking at the attitude question we find that males who generally favor affirmative action for women gender discriminate against males (p-value=0.098), while those who generally oppose do not (p-value=0.490) (column 2). However, the two groups are not significantly different from each other (p-value=0.127). The same is true when controlling for background variables (column 3). Finally, in both samples; females and males, none of the interaction terms are significant when introducing them all in (5) (columns 7 in Tables A.10 and A.11).

the spectators who generally favor affirmative action and for the spectators who generally oppose. The coefficient for $Maleloser_i + Aff.ac_j \times Maleloser_i$ is negative and significant (p<0.000) which means that the spectators who generally favor affirmative action for women discriminate males in the mixed-sex merit environment. This is however, not the case for the spectators who generally oppose affirmative action for women since the coefficient on $Maleloser_i$ is not significantly different from zero (p=0.713) and since the two groups are significantly different from each other (the coefficient on $Aff.ac_j \times Maleloser_i$ is negative and significant, p=0.041). The same result is found in Figure 4 below.

[Figure 4 about here]

Figure 4 shows the mean transfers to the losing worker in the two mixed-sex merit treatments from the second round of data collection. Again, we see that the spectators who generally support affirmative action for women transfer more to the losing worker when the worker is a female rather than a male. We also see that the spectators who generally oppose affirmative action do not gender discriminate in this mixed-sex merit environment.

6 Discussion and conclusion

...There are mixed results concerning the interplay between student and teacher gender. While Lindahl (2015) finds that students graded by same-gender teachers receive lower grades, (Dee, 2007) finds the opposite and Ouazad and Page (2013) that female teachers have a larger gender-bias against males in their grade setting than male teachers. Lavy (2008) again, find that female teachers are more gender-biased against male students in two out of three investigated subject while Terrier (2015) find no effect of teacher gender.

Interessante poenger som kanskje kan vaere nyttig for diskusjonsdelen (sitater fra Bertrand2016):

(s.4-5): Modern social psychologists believe that attitudes can occur in implicit modes and that people can behave in ways that are unrelated or even sometimes opposed to their explicit views or self-interests (Banaji and Greenwald, 1995; Bertrand et al., 2005; Dovidio et al., 1998a,b; Greenwald and Banaji, 1995). Neuroscience studies have shown that (...) Implicit biases are more likely to drive behavior under conditions of ambiguity, high time pressures and cognitive loads, or inattentiveness to the task.

At the same time, evolutionary psychology has stressed the importance of social differentiation and the delineation of clear group boundaries as a way to achieve the benefits of cooperation between human beings without the risk of excessive costs, with group membership and group identity emerging as a form of contingent altruism (Brewer, 1981). While in-group love might not necessarily imply out-group hate, the same factors that make allegiance with group members important provide grounds for antagonism and distrust of outsiders.

the limited information and decision-making model that drives statistical discrimination might be itself endogenous to conscious or unconscious prejudice against the out-group members.

s.6: Perhaps most importantly, whether discrimination is taste-based or statistical, it may ultimately result in genuine difference between groups, through self-fulfilling prophecies. If the stereotypical woman is not good at math, talented girls may become discouraged and ultimately not become good at math. If teachers or employers assume that students of a particular color are less smart, they will invest less in them. Thus, discrimination, whether it is taste-based or statistical, can create or exacerbate existing differences between groups.

s. 17 (fra litteraturen om correspondance studies): Gender There are fewer studies on gender, and discrimination against women at the call-back stage is much less apparent in general. Some studies attempt to show whether the degree (and nature) of discrimination depends on the nature of the profession. Carlsson (2011) sent paired applications for positions of IT professionals, drivers, construction workers, sales assistants, high school teachers, restaurant workers, accountants, cleaners, pre-school teachers, and nurses. Over-all, women are called back slightly more often than men; in male dominated professions, males have a slight (insignificant) advantage. Booth and Leigh (2010) focused on female-dominated professions (waitstaff, data-entry, customer service, and sales jobs) and found a call-back of 1.28 in favor of women.

s. 35 (om Golberg Paradigm experiments): In the typical lab experiment, a group of subjects is asked to review a vignette, describing the behavior of a female or male manager (for example), or witness a confederate (male or female) simulating a leadership situation. The participants are then asked to evaluate the leader's competence, or to say whether they would have liked to have them as leader for a task they may collectively perform. Reviewing a large number of such studies, Eagly, Makhijani, and Klonsky (1992) do not find that, on average, female leaders are evaluated significantly more negatively than male leaders. However, there are in some circumstances where they do find that female leaders are evaluated more negatively: for example, when the leadership was carried out in a masculine style (in particular when the leader was projected to be authoritative). This supports Eagly's hypothesis of "role congruence": what people dislike is when women behave in a non-feminine way. Since strong leaders must be assertive, but women must be demure, it makes it difficult for women to be appreciated as strong leaders. The fact that the circumstances are artificial, and answers have minimal stake associated with them, make those experiments less relevant, on their own, than field-based correspondence tests. Beaman et al. (2009) find that both men and women, but men more than women, tend to discriminate against female leaders (additional results from this study are discussed further in Section 3.1.2).

s.28: The only approach that has been used repeatedly by researchers to try to separate statistical from taste-based discrimination has been to compare differential gaps in outcomes between pairs of minority and non-minority applicants with weaker or stronger productivity attributes on their résumés or applications. As more productivity-relevant information is included on the résumé, average differences in unobservable characteristics between the minority and non-minority applicants are reduced, and statistical discrimination should also be reduced; however, it is clear that this remains a very indirect way to try to isolate taste-based discrimination among employers or landlords.

...To interessante studier vi kan henvise til i diskusjonen (tekst direkte avskrift fra Bertrand2016 (s. 57-58):

"It is, however, not obvious that minority group leaders, or committees that contain such minority leaders, would necessarily favor others from the minority: faced with their own discrimination, they may feel the need to go to lengths to avoid being perceived as biased. In several observational studies, women were not inclined to judge other women more favorably than men. (i fotnote: See Booth and Leigh (2010) for an audit study in Australia, where they find no interaction between the gender on the résumé and the gender of the recruiter. See also Broder (1993) for similar evidence in the context of NSF proposal reviews, and Abrevaya and Hamermesh (2012) for referee reports."

"The empirical evidence of the impact of minority representation on selection committees largely comes from a series of very interesting papers by Bagues, Zinovyeva, and their co-authors. Bagues and Esteve-Volart (2010) examine the impact of the gender composition of the evaluation committee for the entry exam in the Spanish judiciary on the success of women in that exam. A causal study is made possible by the fact that people are randomly assigned to a committee. They find that women are less likely to succeed at the exam when the committee they are assigned to has more women, while the opposite is true for male candidates. Additional evidence in the study suggest that these results might be driven at least in part by the fact that female evaluators tend to overestimate the quality of male candidates."

"Zinovyeva and Bagues (2011) and Bagues, Sylos-Labini, and Zinovyeva (2014) present interesting evidence from randomized academic evaluation committees in Spain and Italy, respectively. In both countries, candidates for promotion appear in front of a centralized committee to be qualified.

Files are assigned to randomly composed committees. In the Spanish case, the authors find no impact of an additional female committee member on the promotion likelihood of female candidates. In the Italian case, they find a negative effect: in a five-member committee, with each additional female member added to the committee, the success rate of female applicants relative to that of male applicants decreases by around two percentage points. Analyzing the voting records, they find both that (1) the same female candidate is scored on more harshly by females than by males, and (2) male committee members grade female candidates more harshly when there are women on the committee, perhaps because they are trying to compensate for a perceived bias in favor of women on such committees (even though in reality the opposite appears to be true given (1))."

"This evidence on academic and recruitment committees is striking and suggest that some type of affirmative action may in fact hurt promising female candidates. It would be interesting to see if it also carries through in other settings, such as management or political decisions."

Og fra s. 82-83:

"In 2010 and 2011, the French government initiated an experiment, which was implemented by the French public employment service. It involved about 1,000 firms in eight local labor markets and it lasted in total for about ten months (Behaghel, Crépon, and Le Barbanchon, 2014). Among volunteer firms, résumés were either transmitted anonymously or non-anonymously. The experiment's main findings can be summarized as follows. First, women benefit from higher callback rates with anonymous job applications - at least if they compete with male applicants for a job. Second, and most interestingly, migrants and residents of deprived neighborhoods suffer from anonymous job applications. (...) Behaghel, Crépon, and Le Barbanchon (2014) explain these surprising results by the self-selection of firms that agreed to participate in the field experiment."

"Another large-scale randomized field experiment took place in Germany in early 2010 (Krause, Rinne, and Zimmermann, 2012a). The publication of a correspondence testing study for Germany (Kaas and Manger, 2012) triggered a lively public debate about discrimination in the hiring decisions of German firms. (...) The characteristics that were made anonymous include the applicant's name and contact details, gender, nationality, date and place of birth, disability, marital status and the applicant's picture.49 Unlike the French study, the authors find that the anonymization leads to less discrimination against minority groups."

References

Almås, Ingvild, Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden (2010). "Fairness and the development of inequality acceptance," *Science*, 328(5982): 1176–1178.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden (2015). "Are Americans more meritocratic and efficiency-seeking than Scandinavians?" mimeo, NHH Norwegian School of Economics.

Autor, David and Melanie Wasserman (2013). "Wayward sons: The emerging gender gap in labor markets and education," Technical report, Third Way.

Balafoutas, Loukas and Matthias Sutter (2012). "Affirmative action policies promote women and do not harm efficiency in the laboratory," *Science*, 335(6068): 579–582.

Bertrand, Marianne (2011). "New perspectives on gender," in Orley C. Ashenfelter and David Card (eds.), "Handbook of Labor Economics," volume 4, chapter 17, Elsevier, pp. 1543–1590.

Bertrand, Marianne and Esther Duflo (2016). "Field Experiments on Discrimination," NBER Working Paper No. 22014.

Bertrand, Marianne and Kevin F. Hallock (2001). "The gender gap in top corporate jobs," *Industrial and Labor Relations Review*, 55(1): 3–21.

Black, Sandra E. and Philip E. Strahan (2001). "Rent-sharing and discrimination in a regulated industry," *The American Economic Review*, 91(4): 814–831.

Bolton, Gary E. and Axel Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, 90(1): 166–193.

Breda, Thomas and Son Thierry Ly (2015). "Professors in Core Science Fields Are Not Always Biased against Women: Evidence from France," *American Economic Journal: Applied Economics*, 7(4): 53–75.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden (2007). "The pluralism of fairness ideals: An experimental approach," *American Economic Review*, 97(3): 818–827.

Cappelen, Alexander W., Karl O. Moene, Erik Ø. Sørensen, and Bertil Tungodden (2013). "Needs versus entitlements: An international fairness experiment," *Journal of the European Economic Association*, 11(3): 574–598.

Cappelen, Alexander W., Erik Ø. Sørensen, and Bertil Tungodden (2010). "Responsibility for what? Fairness and individual responsibility," *European Economic Review*, 54(3): 429–441.

Castillo, Marco, Ragan Petrie, Maximo Torero, and Lise Vesterlund (2013). "Gender differences in bargaining outcomes: A field experiment on discrimination," *Journal of Public Economics*, 99: 35–48.

Charness, Gary and Matthew Rabin (2002). "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, 117(3): 817–869.

Cornwell, Christopher, David B. Mustard, and Jessica Van Parys (2013). "Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school," *Journal of Human Resources*, 48(1): 236–264.

Dee, Thomas S. (2007). “Teachers and the Gender Gaps in Student Achievement,” *The Journal of Human Resources*, XLII(3): 528–554.

DeNavas-Walt, Carmen and Bernadette D. Proctor (2015). “Income and Poverty in the United States: 2014. Current Population Reports,” Technical report, United States Census Bureau.

Dittrich, Dennis A. V., Susanne Büchner, and Micaela M. Kulesz (2015). “Dynamic repeated random dictatorship and gender discrimination,” *Journal of Behavioral and Experimental Economics*, 55: 81–90.

Eagly, Alice H. and Steven J. Karau (2002). “Role Congruity Theory of Prejudice Toward Female Leaders,” *Psychological Review*, 109(3): 573–598.

Engelmann, Dirk and Martin Strobel (2004). “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments,” *American Economic Review*, 94(4): 857–869.

Falch, Torberg and Linn Renée Naper (2013). “Educational evaluation schemes and gender gaps in student achievement,” *Economics of Education Review*, 36: 12–25.

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach (2008). “Egalitarianism in young children,” *Nature*, 454(7208): 1079–1083.

Fehr, Ernst and Klaus M. Schmidt (1999). “A theory of fairness, competition and cooperation,” *Quarterly Journal of Economics*, 114(3): 817–868.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu (2002). “A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition,” *Journal of Personality and Social Psychology*, 82(6): 878–902.

Goldin, Claudia and Cecilia Rouse (2000). “Orchestrating impartiality: The impact of ”blind” auditions on female musicians,” *The American Economic Review*, 90(4): 715–741.

Holzer, Harry and David Neumark (2000). “Assessing Affirmative Action,” *Journal of Economic Literature*, 38(3): 483–568.

Howden, Lindsay M. and Julie A. Meyer (2011). “Age and Sex Composition: 2010. 2010 Census Briefs,” Technical report, United States Census Bureau.

Konow, James (2000). “Fair shares: Accountability and cognitive dissonance in allocation decisions,” *American Economic Review*, 90(4): 1072–1091.

Konow, James, Tatsuyoshi Saijo, and Kenju Akai (2009). “Morals and mores? Experimental evidence on equity and equality,” Mimeo, Loyola Marymount University.

Lavy, Victor (2008). “Do gender stereotypes reduce girls’ or boys’ human capital outcomes? evidence from a natural experiment,” *Journal of Public Economics*, 92(10-11): 2083–2105.

Lindahl, Erica (2015). “Are teacher assessments biased? - evidence from Sweden,” *Education Economics*, 2: 1–15.

List, John A. (2004). “The nature and extent of discrimination in the marketplace: Evidence from the field,” *Quarterly Journal of Economics*, 119(1): 49–89.

Niederle, Muriel, Carmit Segal, and Lise Vesterlund (2013). “How costly is diversity? Affirmative action in light of gender differences in competitiveness,” *Management Science*, 59(1): 1–16.

Ouazad, Amine and Lionel Page (2013). "Students' perceptions of teacher biases: Experimental economics in schools," *Journal of Public Economics*, 105: 116–130.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2014). "How stereotypes impair women's careers in science," *PNAS*, 111(12): 4403–4408.

Robinson, Joseph Paul and Sarah Theule Lubienski (2011). "The Development of Gender Achievement Gaps in Mathematics and Reading During Elementary and Middle School: Examining Direct Cognitive Assessments and Teacher Ratings," *American Educational Research Journal*, 48(2): 268–302.

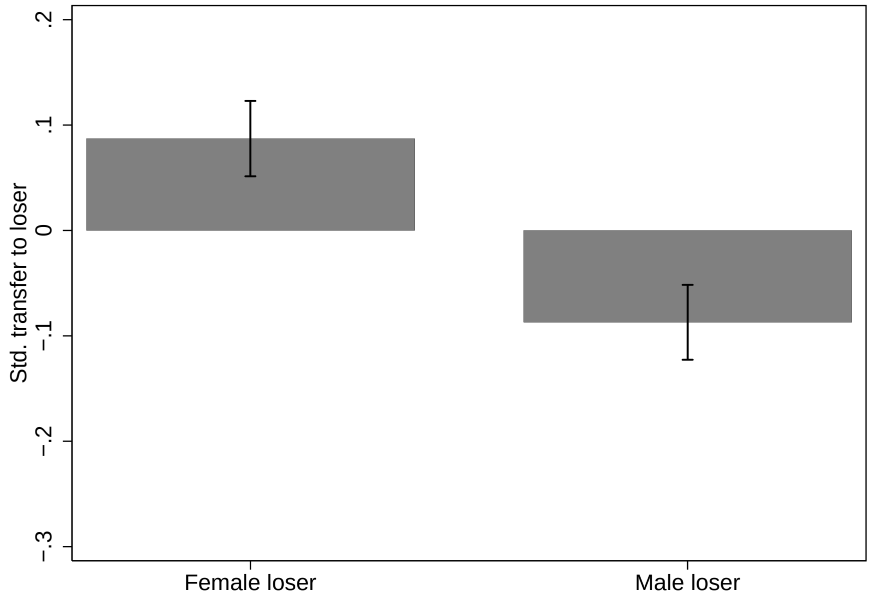
Sharma, Smriti (2015). "Gender and distributional preferences: Experimental evidence from india," *Journal of Economic Psychology*, 50: 113–123.

Sutter, Matthias (2007). "Outcomes versus intentions: On the nature of fair behavior and its development with age," *Journal of Economic Psychology*, 28(1): 69–78.

Terrier, Camille (2015). "Giving a Little Help to Girls? Evidence on Grade Discrimination and its Effect on Students' Achievement," CEP Discussion Paper No 1341.

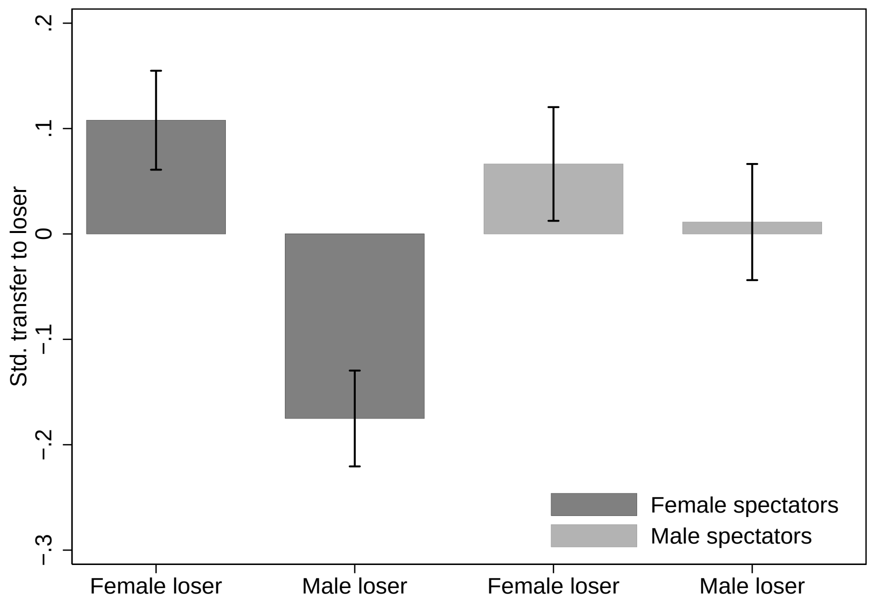
White, Judith B. and Wendi L. Gardner (2009). "Think Women, Think Warm: Stereotype Content Activation in Women with a Salient Gender Identity, Using a Modified Stroop Task," *Sex Roles*, 60: 247–260.

Figure 1: Transfer to loser: Mixed-sex merit



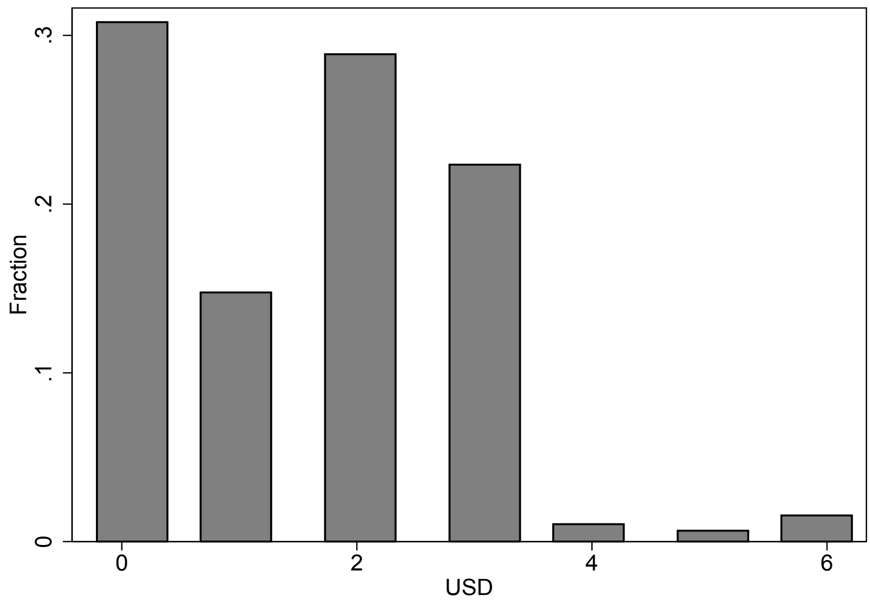
Note: The figure shows the mean transfers to the losing worker in the two mixed-sex merit treatments, T3 (with a female loser) and T4 (with a male loser). The standard errors are indicated by the bars.

Figure 2: Transfer to loser: Mixed-sex merit by gender



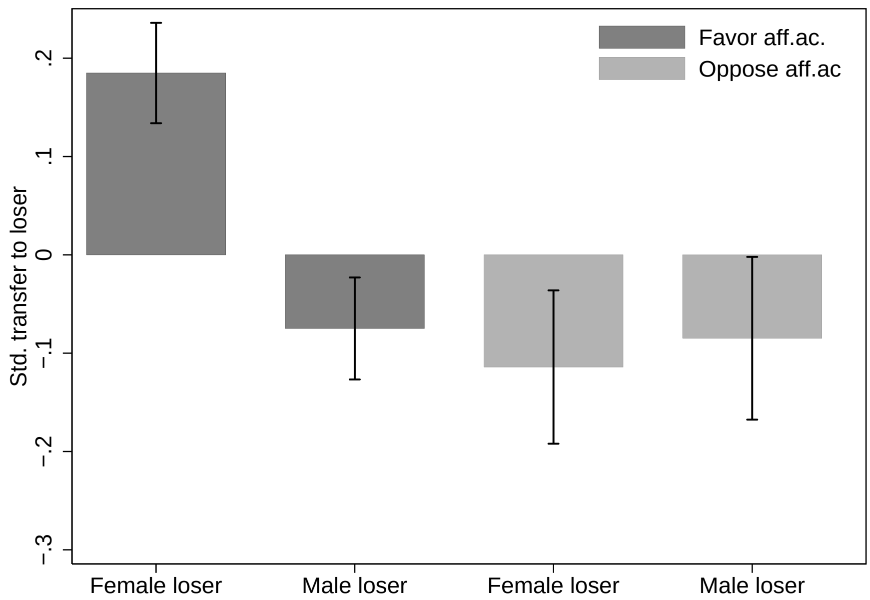
Note: The figure shows the mean transfers to the losing worker in the two mixed-sex merit treatments, T3 (with a female loser) and T4 (with a male loser), split by the gender of the spectators. The standard errors are indicated by the bars.

Figure 3: Transfer to loser



Note: The figure shows the distribution of transfers in USD to the losing worker in all of the treatments seen together (by fraction).

Figure 4: Transfer to loser: Mixed-sex merit by support for affirmative action



Note: The figure shows the mean transfers to the losing worker in the two mixed-sex merit treatments, T3 (with a female loser) and T4 (with a male loser). The sample of spectators is split in two groups; those who generally favor affirmative action for women, and those who generally oppose. The standard errors are indicated by the bars.

Table 1: Descriptive statistics

	Mean	**Male**	**Female**	**Median**	**p10**	**p90**	**sd**
United States							
Income (USD)	68730	72291	65259	55000	19999	125000	40912
Age	41	41	41	41	23	59	13.2
Eduction shares (age at which the person stopped full-time educ.)							
19 years or younger	0.240	0.212	0.266				
20 years or older	0.644	0.674	0.616				
Students	0.116	0.112	0.118				
Share republican	0.337	0.355	0.321				
Share female	0.512	-	-				
Share with child(ren)	0.384	0.363	0.404				

Note: The table displays the descriptive statistics for the background variables of the spectator sample. The income variable is the spectator's combined yearly income in USD (gross income before taxes are deducted) and given in standard categories where we use the mid-point in the category (the categories are "Less than $20.000", "$20.000 to $29.999", "$30.000 to $39.999", "$40.000 to $49.999", "$50.000 to $59.999", "$60.000 to $74.999", "$75.000 to $99.999", "$100.000 to $149.999", "$150.000 or more". A participant is classified as conservative if voting for the Republican party in the US.

Table 2: Sequence of events in the experiment

Stage of experiment
1. **Work stage:** Workers complete an assignment.
2. **Earnings stage:** Workers matched in pairs. Assigned initial earnings according to treatment.
3. **Redistribution stage:** Each spectator decides for one pair of workers whether and how much to redistribute.
4. **Payment stage:** Workers in the pair paid according to the decision of the spectator.

Note: The table provides an overview of the main stages in the experiment.

Table 3: Treatments

	Mixed-sex		Single-sex	
LUCK	T1	T2	T5	T6
	Unlucky: The woman	Unlucky: The man	Two men	Two women
MERIT	T3	T4	T7	T8
	Least productive: The woman	Least productive: The man	Two men	Two women

Note: The table provides an overview of the eight different treatments in the experiment.

Table 4: Regression analysis: Transfer to loser

	Mixed-sex Amount to loser (std)		Mixed-sex Nothing to loser	
	Merit	Merit	Merit	Merit
Male loser	-0.174***	-0.173***	0.073***	0.073***
	(0.050)	(0.050)	(0.024)	(0.024)
Male		0.078		-0.011
		(0.050)		(0.024)
Low age		-0.005		-0.000
		(0.051)		(0.024)
Republican		-0.173***		0.058**
		(0.054)		(0.026)
Low income		-0.011		0.009
		(0.052)		(0.025)
Constant	0.066	0.092	0.335***	0.318***
	(0.051)	(0.068)	(0.024)	(0.033)
Observations	1564	1564	1564	1564
R^2	0.008	0.016	0.006	0.010

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable in columns 1 and 2. The dependent variable in columns 3 and 4 is an indicator variable that takes the value one if the spectator gave nothing to the losing worker. The sample of spectators are from treatments three and four. Male loser is an indicator variable taking the value of one if the spectator is in a treatment where a male lost. Male is an indicator variable that takes the value one if the spectator is male. Low age is an indicator variable that takes the value one if the spectator is below 41 years old (which is the median age in the sample). Republican is an indicator variable that takes the value one if the spectator is republican. Low income is an indicator variable that takes the value one if the spectator has an income below $55,000 (which is the median income per year in the sample). We have also included an indicator variable that takes the value one if the spectator participated in the second round of data collection. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Regression analysis: Transfer to loser (std)

	Mixed-sex		Single-sex			
	Luck		Merit		Luck	
Male loser	-0.033	-0.043	-0.040	-0.031	0.087	0.091
	(0.088)	(0.089)	(0.088)	(0.089)	(0.088)	(0.088)
Male		-0.108		-0.126		-0.020
		(0.089)		(0.088)		(0.088)
Low age		-0.108		0.113		-0.125
		(0.089)		(0.088)		(0.089)
Republican		-0.140		-0.181*		-0.259***
		(0.094)		(0.095)		(0.092)
Low income		-0.027		0.035		-0.033
		(0.092)		(0.092)		(0.091)
Constant	0.017	0.184*	0.020	0.072	-0.044	0.136
	(0.063)	(0.104)	(0.062)	(0.100)	(0.063)	(0.102)
Observations	512	512	513	513	513	513
R^2	0.000	0.010	0.000	0.017	0.002	0.020

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable. The sample of spectators are from treatments one and two in columns 1 and 2, and from treatments seven and eight in columns 3 and 4 and from treatments five and six in columns 5 and 6. Male loser, Male, low age, republican and low income are defined in Table 4. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Heterogeneity analysis: Transfer to loser (std)

	Mixed-sex merit				
	Gender	Age	Politics	Income	All
Male loser	-0.284***	-0.212***	-0.169***	-0.227***	-0.376***
	(0.070)	(0.071)	(0.061)	(0.065)	(0.105)
Male*Male loser	0.227**				0.229**
	(0.100)				(0.101)
Low age*Male loser		0.078			0.071
		(0.101)			(0.101)
Republican*Male loser			-0.012		0.004
			(0.107)		(0.109)
Low income*Male loser				0.133	0.135
				(0.102)	(0.103)
Constant	0.169***	0.134**	0.110*	0.140**	0.201**
	(0.064)	(0.065)	(0.062)	(0.063)	(0.085)
Lincom	-0.056	-0.134*	-0.181**	-0.095	-
(Male loser+interaction)	(0.072)	(0.071)	(0.088)	(0.079)	-
Controls	X	X	X	X	X
Observations	1564	1564	1564	1564	1564
R^2	0.019	0.016	0.015	0.016	0.020

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable using the sample of spectators in treatments three and four. Male loser, male, low age, republican and low income are defined in Table 4. Male*Male loser is an interaction term showing male multiplied with Male loser. Low age*Male loser is an interaction term showing low age multiplied with Male loser. Republican*Male loser is an interaction term showing republican multiplied with Male loser. Low income*Male loser is an interaction term showing low income multiplied with Male loser. Controls include the background variables male, low age, republican, and low income. Column (5) also includes the indicator variable that takes the value one if the spectator participated in the second round of data collection. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Beliefs and attitudes analysis: Transfer to loser (std)

	Mixed-sex merit						
	Aff.ac	Aff.ac	Aff.ac	Overrate	Overrate	Overrate	All
Male loser	-0.175***	0.029	0.032	-0.174***	-0.156**	-0.163**	0.049
	(0.061)	(0.113)	(0.113)	(0.062)	(0.077)	(0.077)	(0.123)
Aff.ac.*Male loser		-0.289**	-0.294**				-0.295**
		(0.135)	(0.135)				(0.135)
Overrate*Male loser					-0.050	-0.030	-0.036
					(0.128)	(0.129)	(0.128)
Aff.ac.	0.154**	0.299***	0.285***				0.284***
	(0.068)	(0.096)	(0.096)				(0.096)
Overrate (math)				0.076	0.100	0.106	0.108
				(0.064)	(0.090)	(0.090)	(0.090)
Constant	-0.011	-0.114	-0.011	0.070	0.061	0.150**	-0.053
	(0.065)	(0.081)	(0.095)	(0.050)	(0.055)	(0.075)	(0.101)
Lincom	-	-0.260***	-0.262***	-	-0.206**	-0.193*	-
(Male loser+interaction)	-	(0.073)	(0.073)	-	(0.103)	(0.103)	-
Controls	-	-	X	-	-	X	X
Observations	1050	1050	1050	1050	1050	1050	1050
R^2	0.013	0.017	0.022	0.009	0.009	0.016	0.024

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable. The sample includes the spectators in treatment three and four from the second round of data collection. Male loser is defined in Table 4. Aff.ac. is an indicator variable taking the value one if the spectator generally favors affirmative action. Aff.ac.*Male loser is an interaction term showing Aff.ac. multiplied with Male loser. Overrate (math) is an indicator variable taking the value one if the spectator overrated males relative to females on belief question 1a. Overrate*Male loser is an indicator term showing Overrate (math) multiplied with Male loser. Controls include the background variables male, low age, republican, and low income defined in Table 4. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Heterogeneity analysis with affirmative action: Transfer to loser (std)

	Mixed-sex merit				
	Gender and Aff.ac.	Age	Politics	Income	All
Male loser	-0.049	-0.212***	-0.169***	-0.227***	-0.063
	(0.133)	(0.071)	(0.061)	(0.065)	(0.164)
Male*Male loser	0.143				0.144
	(0.123)				(0.123)
Aff.ac.*Male loser	-0.277**				-0.316**
	(0.135)				(0.139)
Low age*Male loser		0.078			-0.046
		(0.101)			(0.124)
Republican*Male loser			-0.012		-0.075
			(0.107)		(0.134)
Low income*Male loser				0.133	0.211*
				(0.102)	(0.126)
Constant	-0.027	0.134**	0.110*	0.140**	-0.018
	(0.106)	(0.065)	(0.062)	(0.063)	(0.117)
Lincom	0.094	-0.134*	-0.181**	-0.095	-
(Male loser+interaction)	(0.125)	(0.071)	(0.088)	(0.079)	-
Lincom	-0.326***	-	-	-	-
(Male loser+Aff.ac.*Male loser)	(0.092)	-	-	-	-
Controls	X	X	X	X	X
Observations	1050	1564	1564	1564	1050
R^2	0.026	0.016	0.015	0.016	0.029

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable using the sample of spectators in treatments three and four. Male loser, male, low age, republican and low income are defined in Table 4. Male*Male loser, Low age*Male loser, Republican*Male loser and Low income*Male loser are defined in Table 6. Aff.ac. and Aff.ac.*Male loser are defined in Table 7. Controls include the background variables male, low age, republican, and low income as well as Aff.ac. Column (5) also includes the indicator variable that takes the value one if the spectator participated in the second round of data collection. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix

A.1 Treatment instructions

Treatment 1: Luck, mixed-sex, unlucky woman

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. A few days ago two workers were recruited via an online labor market to conduct an assignment. They were both from the US; a man and a woman of the same age.

They were each paid a participation compensation of 2 USD regardless of what they would end up being paid for the assignment. After completing the assignment, they were told that their earnings from the assignment would be determined by a lottery. The worker winning the lottery would earn 6 USD for the assignment and the other worker would earn nothing for the assignment. They were not informed about the outcome of the lottery. However, they were told that a third person would be informed about the assignment and the outcome of the lottery. They were also told that this third person would be given the opportunity to redistribute the earnings and thus determine how much they were paid for the assignment.

You are the third person and we now want you to choose whether to redistribute the earnings for the assignment between the two workers. Your decision is completely anonymous. The workers will receive the payment that you choose for the assignment within a few days, but will not receive any further information.

The man was lucky, won the lottery and earned 6 USD for the assignment. The woman was unlucky and earned nothing for the assignment.

Please state which of the following alternatives you choose:

I do not redistribute:

- *The lucky worker is paid 6 USD and the unlucky worker is paid 0 USD.*

I do redistribute:

- *The lucky worker is paid 5 USD and the unlucky worker is paid 1 USD.*
- *The lucky worker is paid 4 USD and the unlucky worker is paid 2 USD.*
- *The lucky worker is paid 3 USD and the unlucky worker is paid 3 USD.*
- *The lucky worker is paid 2 USD and the unlucky worker is paid 4 USD.*
- *The lucky worker is paid 1 USD and the unlucky worker is paid 5 USD.*
- *The lucky worker is paid 0 USD and the unlucky worker is paid 6 USD.*

Treatment 8: Merit, single-sex, two women

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. A few days ago two workers were recruited via an online labor market to conduct an assignment. They were both from the US; two women of the same age.

They were each paid a participation compensation of 2 USD regardless of what they would end up being paid for the assignment. After completing the assignment, they were told that their earnings from the assignment would be determined by their productivity. The most productive worker would earn 6 USD for the assignment and the other worker would earn nothing for the assignment. They were not informed about who was the most productive worker. However, they were told that a third person would be informed about the assignment and who was the most productive worker. They were also told that this third person would be given the opportunity to redistribute the earnings and thus determine how much they were paid for the assignment.

You are the third person and we now want you to choose whether to redistribute the earnings for the assignment between the two workers. Your decision is completely anonymous. The workers will receive the payment that you choose for the assignment within a few days, but will not receive any further information.

One of the women was most productive and earned 6 USD for the assignment. The other woman was least productive and earned nothing for the assignment.
Please state which of the following alternatives you choose:

I do not redistribute:

- *The most productive worker is paid 6 USD and the least productive worker is paid 0 USD.*

I do redistribute:

- *The most productive worker is paid 5 USD and the least productive worker is paid 1 USD.*
- *The most productive worker is paid 4 USD and the least productive worker is paid 2 USD.*
- *The most productive worker is paid 3 USD and the least productive worker is paid 3 USD.*
- *The most productive worker is paid 2 USD and the least productive worker is paid 4 USD.*
- *The most productive worker is paid 1 USD and the least productive worker is paid 5 USD.*
- *The most productive worker is paid 0 USD and the least productive worker is paid 6 USD.*

A.2 Background questions

In addition, the spectators will answer the following set of background questions:

- (...) we would be grateful if you could type in your actual age below?
- Are you?
  - Male
  - Female
- In which region do you live? (State in the United States)
- What is your household's combined yearly income (gross income before taxes are deducted)?
  - Less than $20.000
  - $20.000 - $29.999

- $30.000 - $39.999
- $40.000 - $49.999
- $50.000 - $59.999
- $60.000 - $74.999
- $75.000 - $99.999
- $100.000 - $149.999
- $150.000 or more

• Which political party would you vote for if there was an election tomorrow?
  - Republican
  - Democratic
  - Other

• Including yourself, how many individuals aged (16 or 18) or over live in your household?
  - 1/2/3/4/5/6/7/8 or more.

• How many children aged (17 or 15) or under live in your household?
  - 1/2/3/4/5/6/7/8 or more.

• Which of the following best describes you?
  - I am solely responsible for all/most of the grocery shopping in my household.
  - I am jointly responsible for all/most of the grocery shopping in my household.
  - Somebody else in the household takes care of all/most of the grocery shopping.

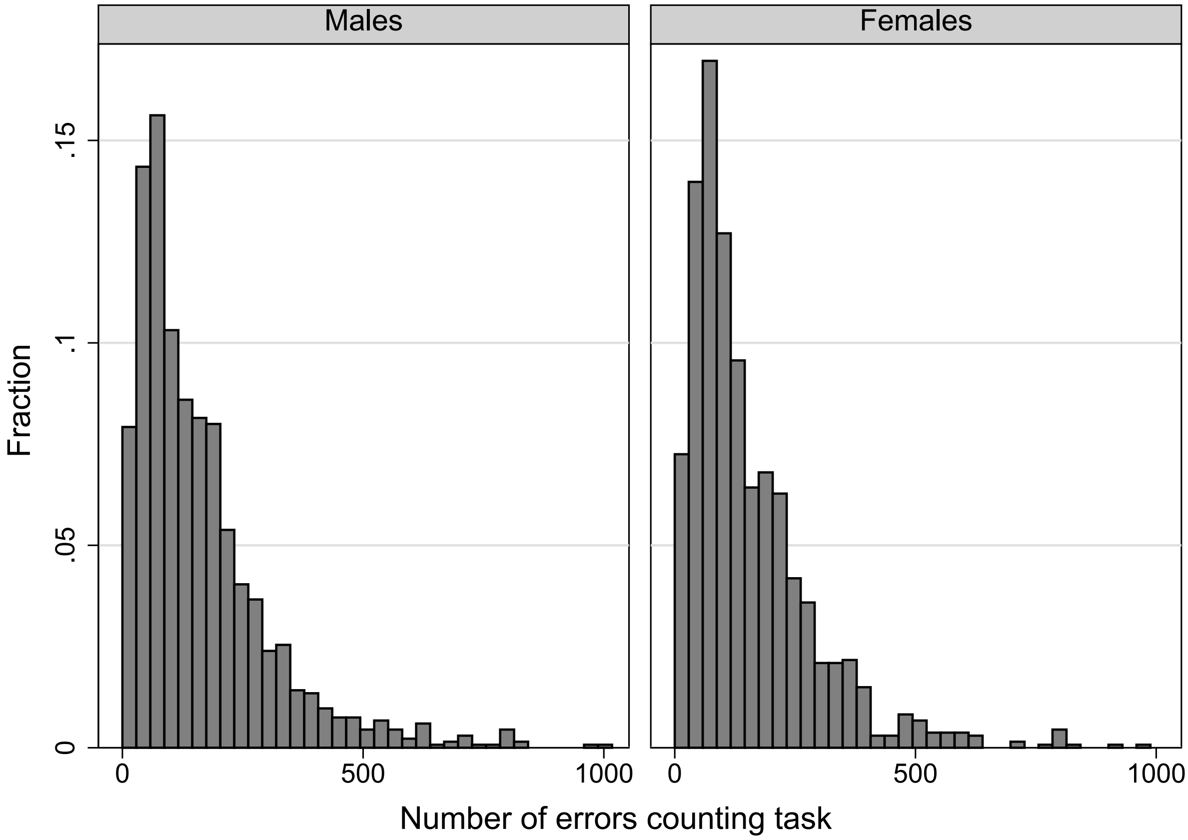
• Which of the following groups does the Chief Income Earner in your household belong to...[19]
  - Semi or unskilled manual worker (e.g. Manual jobs that require no special training or qualifications; Manual workers, Apprentices to be skilled trades, Caretaker, Cleaner, Nursery School Assistant, Park keeper, Non-HGV driver, Shop assistant etc.).
  - Skilled manual worker (e.g. Skilled Bricklayer, Carpenter, Plumber, Painter, Bus/Ambulance Driver, HGV driver, Unqualified assistant teacher, AA patrolman, pub/bar worker, etc.).
  - Supervisory or clerical/ Junior managerial/ Professional/ Administrator (e.g. Office worker, Student Doctor, Foreman with 25+ employees, Sales person, Student Teachers etc.).
  - Intermediate managerial/ Professional/ Administrative (e.g. Newly qualified (under 3 years) doctor, Solicitor, Board director small organisation, Middle manager in large organisation, Principle officer in civil service/local government etc.).
  - Higher managerial/ Professional/ Administrative (e.g. Established doctor, Solicitor, Board Director in large organisation (200+ employees, Top level civil servant/Public service employee), Headmaster/Mistress, etc.).
  - Student
  - Retired or living on pension only.

[19]Additional explanations for the participants: The person in the household with the largest income is the Chief Income Earner, however this income is obtained. If the Chief Income Earner is retired and has an occupational pension, please select according to theÂ previous occupation. If the Chief Income Earner is not in paid employment and has been out of work for less than 6 months, please selectÂ according to previous occupation.

  - Unemployed (for over 6 months) or not working due to long term sickness.
- What is your current employment status?
  - Employed full time (30+ hours per week).
  - Employed part time (less than 30 hours per week).
  - Self-employed
  - Retired/Unable to work/Disabled.
  - Still at school.
  - In full time higher education.
  - Unemployed and seeking work.
  - Not working and not seeking work.
  - Prefer not to say.
- How old were you when you stopped full-time education?
  - 16 years or younger.
  - 17-19 years.
  - 20 years or older.
  - Still studying.
- What is your domestic status? Which of these best describes your situation?
  - Married/living with partner.
  - Never married (single).
  - Divorced/widowed.
  - Living with parents.
  - Domestic partner/living with other adults.
  - Prefer not to state/other.

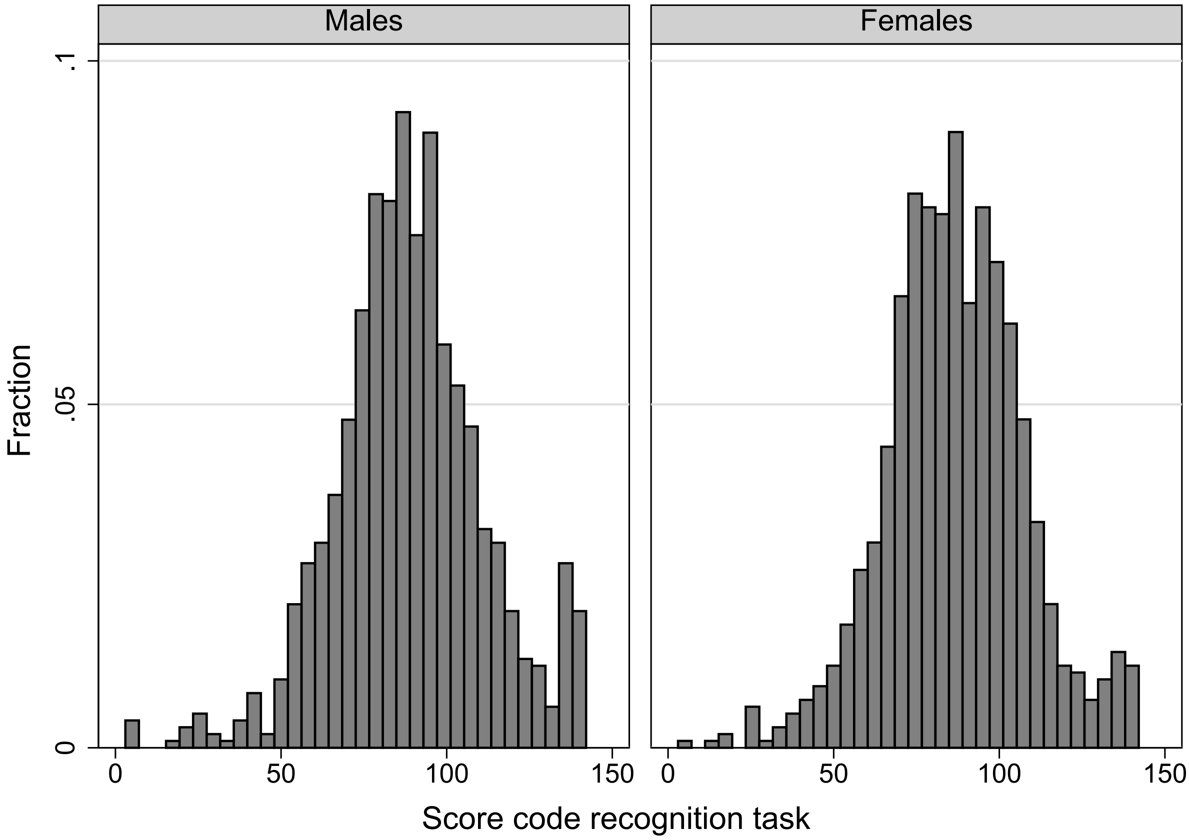
A.3 Figures and tables

Figure A.1: Task productivity, counting task



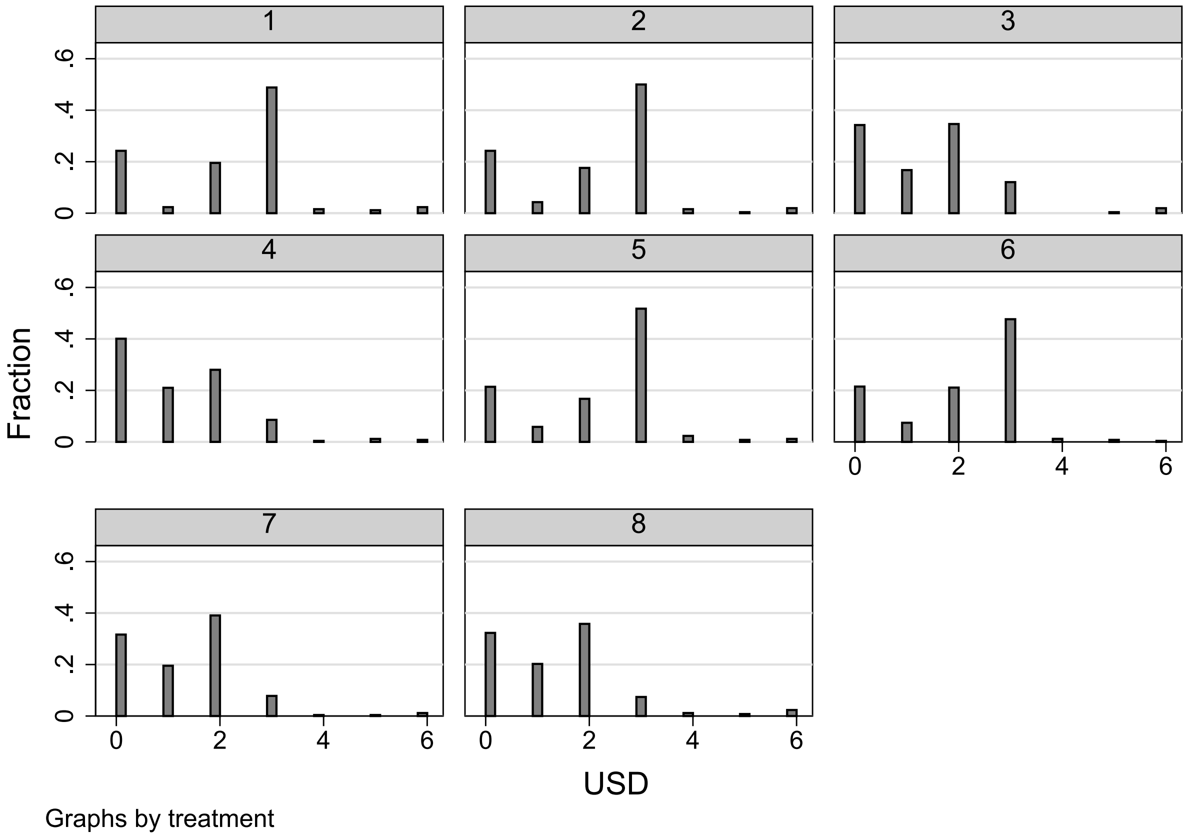
Note: The figure shows the distribution of the worker performance on the counting task. The performance is measured by the number of errors. Five outliers out of 2683 observations have been excluded (those above 1000 errors). The left figure shows the performance distribution among male workers, and the right figure shows the performance distribution among female workers.

Figure A.2: Task productivity, code recognition task



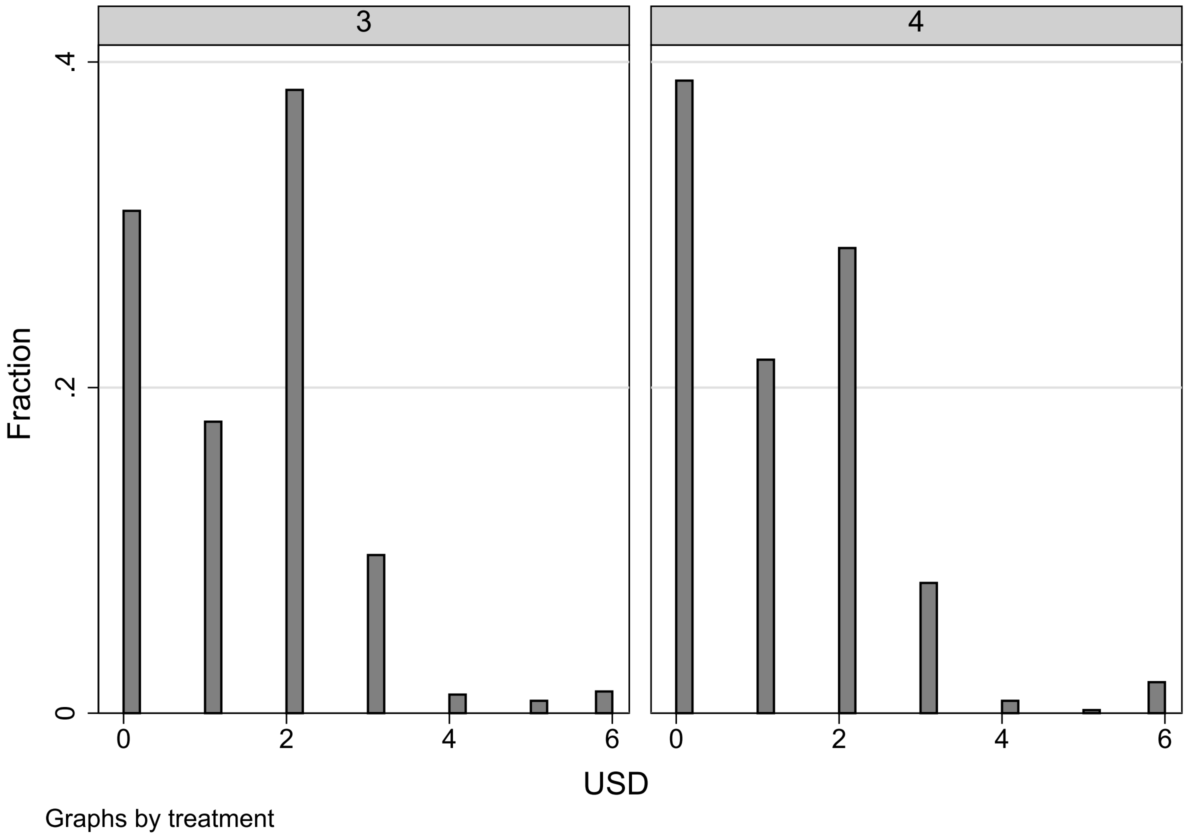
Note: The figure shows the distribution of the worker performance on the code recognition task. The performance is measured by the score of each worker. Two outliers out of 2011 observations have been excluded (those with a score below 0). The left figure shows the performance distribution among male workers, and the right figure shows the performance distribution among female workers.

Figure A.3: Distribution of transfers to loser by treatment, round 1



Note: The figure shows the distribution of transfers in USD to the losing worker in each of the treatments (by fraction). The sample is made up of spectators from the first round of data collection, and figure 1-8 represents Treatment 1- Treatment 8, respectively.

Figure A.4: Distribution of transfers to loser by treatment, round 2



Note: The figure shows the distribution of transfers in USD to the losing worker in each of the treatments (by fraction). The sample is made up of spectators from the second round of data collection, and figure 3 and 4 represents Treatment 3 and Treatment 4, respectively.

Table A.1: Balance test

	(1) Income	(2) Age	(3) Education	(4) Republican	(5) Female	(6) Share with child(ren)
T2_dummy	7557.009**	1.406	0.031	-0.039	-0.031	-0.008
	(3684.318)	(1.167)	(0.055)	(0.042)	(0.044)	(0.043)
T3_dummy	-538.868	0.075	-0.009	-0.026	-0.020	0.032
	(3356.926)	(1.063)	(0.050)	(0.038)	(0.040)	(0.039)
T4_dummy	1916.175	0.995	-0.018	-0.031	0.008	-0.004
	(3355.574)	(1.063)	(0.050)	(0.038)	(0.040)	(0.039)
T5_dummy	118.205	2.228*	-0.038	0.049	-0.018	0.041
	(3684.318)	(1.165)	(0.055)	(0.042)	(0.044)	(0.043)
T6_dummy	2645.751	-0.496	0.074	0.004	-0.020	0.004
	(3691.922)	(1.167)	(0.055)	(0.042)	(0.044)	(0.043)
T7_dummy	1822.693	1.727	-0.070	-0.020	-0.047	-0.000
	(3691.922)	(1.167)	(0.055)	(0.042)	(0.044)	(0.043)
T8_dummy	2538.099	0.943	0.001	-0.036	-0.037	0.014
	(3711.419)	(1.165)	(0.055)	(0.042)	(0.044)	(0.043)
Included control:						
Round dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2954	3102	3102	3102	3102	3102
R^2	0.005	0.004	0.004	0.002	0.001	0.001
Prob > F	0.3692	0.1718	0.2664	0.4264	0.8637	0.7981

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table reports OLS regressions with the different background variables as outcome variables. For each background variable, we have reported the p-value of the joint F-test testing whether the eight treatments are significantly different from each other with respect to that background variable. T2-T8 are indicator variables indicating whether the spectator belongs to treatment 2, 3,... or 8, respectively (yielding the value one) or if they do not (yielding the value zero). The reference category across all regressions is treatment 1. The dependent variable in column 1 is income. Income is the spectator's reported combined, yearly household yearly income (gross income before taxes are deducted). The nine different income categories used are specified in Appendix A.1. The dependent variable in column 2 is age, which reports the spectator's age in number of years. The dependent variable in column 3 is education. Education is a variable where the spectators answered when they stopped full-time education with four different categories specified in Appendix A.1. The dependent variable in column 4 is republican, an indicator variable that takes the value one if the spectator is republican and zero otherwise. The dependent variable in column 5 is female, an indicator variable that takes the value one if the spectator is a female and zero otherwise. The dependent variable in column 6 is share with child(ren), a variable that takes the value one if the spectator has one or more children in the household and zero otherwise. All regressions include the indicator variable that takes the value one if the spectator participated in the second round of data collection. Standard errors in parentheses, where * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.2: Regression analysis by round: Transfer to loser

	Mixed-sex Round 1		Mixed-sex Round 2	
	Merit	Merit	Merit	Merit
Male loser	-0.170*	-0.169*	-0.176***	-0.175***
	(0.088)	(0.088)	(0.062)	(0.061)
Male		0.052		0.093
		(0.088)		(0.061)
Low age		0.127		-0.069
		(0.089)		(0.062)
Republican		-0.207**		-0.161**
		(0.095)		(0.066)
Low income		-0.018		-0.017
		(0.092)		(0.063)
Constant	0.064	0.049	0.098**	0.147**
	(0.062)	(0.103)	(0.044)	(0.072)
Observations	514	514	1050	1050
R^2	0.007	0.022	0.008	0.016

Note: The table reports OLS regressions with the standardized amount given to the losing worker (based on the two mixed-sex merit treatments). The sample of spectators are from treatments three and four, but only from data collection round 1 in columns 1 and 2, and only from round 2 in columns 3 and 4. Male loser, male, low age, republican and low income are defined in Table 4. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Joint F-test luck- and single-sex treatments

	Transfer to loser (std)	Transfer to loser (std)
Male loser	-0.035	-0.042
	(0.085)	(0.085)
Luck single-sex	0.005	0.013
	(0.085)	(0.085)
Merit single-sex	-0.608***	-0.612***
	(0.085)	(0.085)
Luck single-sex*Male loser	-0.049	-0.046
	(0.121)	(0.120)
Merit single-sex*Male loser	0.071	0.077
	(0.121)	(0.120)
Constant	0.215***	0.345***
	(0.060)	(0.076)
Lincom	0.180***	0.303***
(Constant+Male loser)	(0.060)	(0.074)
Lincom	-0.043	-0.034
(Luck single-sex+Luck single-sex*Male loser)	(0.085)	(0.085)
Lincom	-0.537***	-0.535***
(Merit single-sex+Merit single-sex*Male loser)	(0.085)	(0.085)
Controls	-	X
Observations	1538	1538
R^2	0.071	0.081
Prob > F	0.5923	0.5623

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable (standardized based on the luck- and the single-sex treatments). The sample includes all spectators, except for those in treatment 3 and 4 (the mixed-sex merit treatments). Male loser is defined in Table 4. Luck single-sex is an indicator variable taking the value one if the spectator belongs to either of the two single-sex luck treatments. Merit single-sex is an indicator variable taking the value one if the spectator belongs to either of the two single-sex merit treatments. Luck single-sex*Male loser is an interaction term showing Luck single-sex multiplied with Male loser. Merit single-sex*Male loser is an interaction term showing Merit single-sex multiplied with Male loser. We report the p-value of the joint F-test testing whether the Male loser, Luck single-sex*Male loser and Merit single-sex*Male loser are significantly different from each other with respect to the standardized amount given to the losing worker. Controls include the background variables male, low age, republican, and low income defined in Table 4. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Heterogeneity analysis: Transfer to loser (std)

	Mixed-sex luck				
	Gender	Age	Politics	Income	All
Male loser	0.029	0.063	-0.045	-0.194*	-0.035
	(0.123)	(0.125)	(0.108)	(0.114)	(0.183)
Male*Male loser	-0.149				-0.109
	(0.178)				(0.178)
Low age*Male loser		-0.216			-0.252
		(0.178)			(0.178)
Republican*Male loser			0.007		0.036
			(0.188)		(0.189)
Low income*Male loser				0.380**	0.396**
				(0.181)	(0.183)
Constant	0.146	0.125	0.185*	0.263**	0.174
	(0.113)	(0.115)	(0.108)	(0.110)	(0.133)
Lincom	-0.120	-0.153	-0.039	0.186	-
(Male loser+interaction)	(0.128)	(0.127)	(0.155)	(0.140)	-
Controls	X	X	X	X	X
Observations	512	512	512	512	512
R^2	0.012	0.013	0.010	0.019	0.024

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable using the sample of spectators in treatments one and two. Male loser, male, low age, republican and low income are defined in Table 4. Male*Male loser, Low age*Male loser, Republican*Male loser and Low income*Male are defined in Table 6. Controls include the background variables male, low age, republican, and low income. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Heterogeneity analysis: Transfer to loser (std)

	Single-sex luck				
	Gender	Age	Politics	Income	All
Female single-sex	-0.109	-0.001	-0.062	0.071	0.194
	(0.123)	(0.124)	(0.112)	(0.111)	(0.183)
Male*Female single-sex	0.036				0.032
	(0.177)				(0.177)
Low age*Female single-sex		-0.182			-0.190
		(0.177)			(0.178)
Republican*Female single-sex			-0.078		-0.120
			(0.183)		(0.184)
Low income*Female single-sex				-0.428**	-0.427**
				(0.181)	(0.181)
Constant	0.236**	0.182*	0.208*	0.142	0.075
	(0.110)	(0.110)	(0.110)	(0.107)	(0.133)
Lincom	-0.072	-0.184	-0.139	-0.357**	-
(Female single-sex+interaction)	(0.127)	(0.126)	(0.144)	(0.142)	-
Controls	X	X	X	X	X
Observations	513	513	513	513	513
R^2	0.020	0.022	0.020	0.030	0.033

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable using the sample of spectators in treatments five and six (the single-sex luck treatments). Female single-sex is an indicator variable that is equal to one if the spectator is in a treatment of two females. Male, low age, republican and low income are defined in Table 4. Male*Female single-sex is an interaction term showing male multiplied with Female single-sex. Low age*Female single-sex is an interaction term showing low age multiplied with Female single-sex. Republican*Female single-sex is an interaction term showing republican multiplied with Female single-sex. Low income*Female single-sex is an interaction term showing low income multiplied with Female single-sex. Controls include the background variables male, low age, republican, and low income. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Heterogeneity analysis: Transfer to loser (std)

	Single-sex merit				
	Gender	Age	Politics	Income	All
Female single-sex	0.101	-0.094	0.074	0.108	0.109
	(0.126)	(0.122)	(0.108)	(0.111)	(0.181)
Male*Female single-sex	-0.138				-0.147
	(0.176)				(0.177)
Low age*Female single-sex		0.265			0.257
		(0.177)			(0.177)
Republican*Female single-sex			-0.130		-0.130
			(0.189)		(0.190)
Low income*Female single-sex				-0.209	-0.227
				(0.184)	(0.185)
Constant	0.005	0.099	0.016	-0.002	-0.011
	(0.111)	(0.108)	(0.107)	(0.108)	(0.130)
Lincom	-0.036	0.170	-0.056	-0.102	-
(Female single-sex+interaction)	(0.124)	(0.128)	(0.155)	(0.147)	-
Controls	X	X	X	X	X
Observations	513	513	513	513	513
R^2	0.018	0.021	0.017	0.019	0.026

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable using the sample of spectators in treatments seven and eight (the single-sex merit treatments). Female single-sex is an indicator variable that is equal to one if the spectator is in a treatment of two females. Male, low age, republican and low income are defined in Table 4. Male*Female single-sex is an interaction term showing male multiplied with Female single-sex. Low age*Female single-sex is an interaction term showing low age multiplied with Female single-sex. Republican*Female single-sex is an interaction term showing republican multiplied with Female single-sex. Low income*Female single-sex is an interaction term showing low income multiplied with Female single-sex. Controls include the background variables male, low age, republican, and low income. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Heterogeneity analysis: Transfer to loser (std)

	Mixed-sex merit				
	Gender	Age	Politics	Income	All
Male loser	-0.284***	-0.212***	-0.169***	-0.227***	-0.376***
	(0.070)	(0.071)	(0.061)	(0.065)	(0.105)
Male*Male loser	0.227**				0.229**
	(0.100)				(0.101)
Low age*Male loser		0.078			0.071
		(0.101)			(0.101)
Republican*Male loser			-0.012		0.004
			(0.107)		(0.109)
Low income*Male loser				0.133	0.135
				(0.102)	(0.103)
Male	-0.035	0.078	0.079	0.080	-0.036
	(0.071)	(0.050)	(0.050)	(0.050)	(0.071)
Low age	-0.007	-0.044	-0.005	-0.007	-0.044
	(0.051)	(0.072)	(0.051)	(0.051)	(0.072)
Republican	-0.172***	-0.175***	-0.167**	-0.173***	-0.177**
	(0.054)	(0.054)	(0.076)	(0.054)	(0.077)
Low income	-0.009	-0.012	-0.010	-0.076	-0.078
	(0.052)	(0.052)	(0.052)	(0.072)	(0.073)
Constant	0.169***	0.134**	0.110*	0.140**	0.201**
	(0.064)	(0.065)	(0.062)	(0.063)	(0.085)
Lincom	-0.056	-0.134*	-0.181**	-0.095	-
(Male loser+interaction)	(0.072)	(0.071)	(0.088)	(0.079)	-
Observations	1564	1564	1564	1564	1564
R^2	0.019	0.016	0.015	0.016	0.020

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable using the sample of spectators in treatments three and four. Male loser, male, low age, republican and low income are defined in Table 4. Male*Male loser, Low age*Male loser, Republican*Male loser and Low income*Male are defined in Table 6. Column (5) also includes the indicator variable that takes the value one if the spectator participated in the second round of data collection. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Heterogeneity analysis including education: Transfer to loser (std)

	Mixed-sex merit					
	Gender	Age	Politics	Income	Educ.	All
Male loser	-0.285***	-0.210***	-0.168***	-0.228***	-0.176***	-0.376***
	(0.070)	(0.071)	(0.062)	(0.065)	(0.058)	(0.108)
Male*Male loser	0.232**					0.234**
	(0.101)					(0.101)
Low age*Male loser		0.076				0.067
		(0.101)				(0.101)
Republican*Male loser			-0.013			0.003
			(0.107)			(0.109)
Low income*Male loser				0.135		0.138
				(0.102)		(0.105)
Low educ.*Male loser					0.012	-0.003
					(0.119)	(0.121)
Constant	0.159**	0.124*	0.100	0.131**	0.104*	0.188**
	(0.065)	(0.066)	(0.063)	(0.064)	(0.062)	(0.086)
Lincom	-0.053	-0.135*	-0.182**	-0.093	-0.163	-
(Male loser +interaction)	(0.072)	(0.071)	(0.088)	(0.079)	(0.104)	-
Controls	X	X	X	X	X	X
Observations	1564	1564	1564	1564	1564	1564
R^2	0.019	0.016	0.016	0.017	0.016	0.021

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable using the sample of spectators in treatments three and four. Male loser, male, low age, republican and low income are defined in Table 4. Male*Male loser, Low age*Male, Republican*Male loser and Low income*Male are defined in Table 6. Low educ. is an indicator variable that takes the value one if the spectator stopped his/her education at age 19 or younger. Low educ.*Male loser is an interaction term showing low educ. multiplied with Male loser. Controls include the background variables male, low age, republican, low income and low educ. Column (6) also includes the indicator variable that takes the value one if the spectator participated in the second round of data collection. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Regression analysis by gender: Transfer to loser

	Mixed-sex Female spectators		Mixed-sex Male spectators	
	Merit	Merit	Merit	Merit
Male loser	-0.283***	-0.284***	-0.055	-0.057
	(0.065)	(0.065)	(0.077)	(0.077)
Low age		-0.031		0.019
		(0.065)		(0.078)
Republican		-0.201***		-0.142*
		(0.071)		(0.083)
Low income		-0.035		0.016
		(0.066)		(0.080)
Constant	0.103	0.196**	0.066	0.068
	(0.066)	(0.084)	(0.054)	(0.097)
Observations	804	804	760	760
R^2	0.023	0.033	0.001	0.006

Note: The table reports OLS regressions with the standardized amount given to the losing worker (standardized based on both male and females spectators in the mixed-sex merit treatments). The sample of spectators are from treatments three and four, but only female spectators in columns 1 and 2, and only male spectators in columns 3 and 4. Male loser, low age, republican and low income are defined in Table 4. All regressions also include an indicator variable that takes the value one if the spectator participated in the second round of data collection. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Beliefs and attitudes analysis for female spectators: Transfer to loser (std)

	Mixed-sex merit						
	Aff.ac	Aff.ac	Aff.ac	Overrate	Overrate	Overrate	All
Male loser	-0.239***	-0.072	-0.062	-0.237***	-0.219**	-0.226**	-0.043
	(0.075)	(0.151)	(0.150)	(0.075)	(0.093)	(0.093)	(0.160)
Aff.ac.	0.072	0.184	0.170				0.169
	(0.087)	(0.124)	(0.124)				(0.125)
Aff.ac.*Male loser		-0.222	-0.239				-0.242
		(0.174)	(0.173)				(0.173)
Overrate (math)				0.060	0.084	0.100	0.101
				(0.079)	(0.110)	(0.110)	(0.110)
Overrate*Male loser					-0.050	-0.035	-0.037
					(0.157)	(0.157)	(0.157)
Constant	-0.177**	-0.261**	-0.143	-0.145**	-0.154**	-0.054	-0.180
	(0.084)	(0.107)	(0.120)	(0.061)	(0.067)	(0.090)	(0.128)
Lincom	-	-0.294***	-0.301***	-	-0.269**	-0.261**	-
(Male loser+interaction)	-	(0.086)	(0.086)	-	(0.127)	(0.126)	-
Controls	-	-	X	-	-	X	X
Observations	532	532	532	532	532	532	532
R^2	0.020	0.023	0.036	0.020	0.020	0.034	0.038

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable. The sample includes the female spectators in treatment three and four from the second round of data collection. Male loser, Aff.ac., Aff.ac.*Male loser, Overrate (math) and Overrate*Male loser are defined in Table 7. Controls include the background variables low age, republican, and low income defined in Table 4. Standard errors in parentheses, where * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.11: Beliefs and attitudes analysis for female spectators: Transfer to loser (std)

	Mixed-sex merit						
	Aff.ac	Aff.ac	Aff.ac	Overrate	Overrate	Overrate	All
Male loser	-0.082	0.102	0.100	-0.082	-0.064	-0.068	0.121
	(0.086)	(0.148)	(0.149)	(0.087)	(0.109)	(0.109)	(0.164)
Aff.ac.	0.223**	0.362***	0.355***				0.355***
	(0.091)	(0.128)	(0.130)				(0.130)
Aff.ac.*Male loser		-0.278	-0.278				-0.279
		(0.182)	(0.183)				(0.183)
Overrate (math)				0.077	0.102	0.104	0.106
				(0.090)	(0.126)	(0.126)	(0.126)
Overrate*Male loser					-0.050	-0.038	-0.046
					(0.180)	(0.181)	(0.180)
Constant	-0.271***	-0.363***	-0.298**	-0.151**	-0.160**	-0.099	-0.341**
	(0.086)	(0.105)	(0.129)	(0.070)	(0.077)	(0.108)	(0.140)
Lincom	-	-0.175*	-0.177*	-	-0.114	-0.107	-
(Male loser+interaction)	-	(0.106)	(0.106)	-	(0.144)	(0.144)	-
Controls	-	-	X	-	-	X	X
Observations	518	518	518	518	518	518	518
R^2	0.013	0.018	0.020	0.003	0.003	0.006	0.021

Note: The table reports OLS regressions with the standardized amount given to the losing worker as the outcome variable. The sample includes the male spectators in treatment three and four from the second round of data collection. Male loser, Aff.ac., Aff.ac.*Male loser, Overrate (math) and Overrate*Male loser are defined in Table 7. Controls include the background variables low age, republican, and low income defined in Table 4. Standard errors in parentheses, where * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.