European Economic Review 60 (2013) 144–169

Contents lists available at SciVerse ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Updating, self-confidence, and discrimination

Konstanze Albrecht [a,b,*], Emma von Essen [c], Juliane Parys [d], Nora Szech [e,*]

[a] Department of Economics, Institute for Empirical Research in Economics, University of Bonn, Lennéstr. 43, 53113 Bonn, Germany
[b] Center for Economics and Neuroscience, University of Bonn, Nachtigallenweg 86, 53127 Bonn, Germany
[c] Department of Economics, Stockholm University, SE-10691, Stockholm, Sweden
[d] Department of Economics, University of Bonn, Kaiserstr. 1, 53113 Bonn, Germany
[e] Department of Economics, Chair of Economics – Industrial Organization, University of Bamberg, Feldkirchenstr. 21, 96052 Bamberg, Germany

ARTICLE INFO

Article history:
Received 28 February 2012
Accepted 8 February 2013
Available online 18 February 2013

JEL classification:
D81
C91
J16

Keywords:
Updating
Conservatism
Gender
Discrimination
Self-confidence

ABSTRACT

In this laboratory experiment, we show that people incorporate irrelevant group information when evaluating others. Individuals from groups that perform badly on average receive low evaluations, even when it is known that the individuals themselves perform well. This group-bias occurs both in a gendered setup, where women form the worse performing group, and in a non-gendered setup.

The type of discrimination that we identify is neither taste-based nor statistical; it is rather due to conservatism in updating beliefs, and is even more pronounced among women. Furthermore, self-confident men overvalue male performers. When our data is used to simulate a job promotion ladder, we observe that women are driven out quickly.

1. Introduction

Consider two individuals who perform equally well. Is their performance evaluated the same way? In a laboratory experiment, we find that this is not the case if one individual belongs to a less favorable group. We observe this group-bias in evaluations both in a gendered setup, where women form the worse performing group on average, and in a non-gendered setup. Discrimination occurs even though the evaluator knows that group belonging is completely irrelevant. Hence, the discrimination we identify has nothing to do with statistical discrimination. Also, no interaction occurs between evaluator and performer, and thus there is no room for taste-based discrimination; the discrimination we find is rather due to conservatism in updating beliefs. This conservatism is even more pronounced among female evaluators. Furthermore, in the gendered setup, we find that self-confident male evaluators overvalue male performers.

In our study, the information on group belonging is given as the prior (or base-rate information), and the information on the individual as the new information. We hence analyze a simple updating problem. In the economic and psychological literature on updating, there is broad evidence that people do not update according to Bayes' rule; instead, they either are too

* Corresponding authors.
E-mail addresses: konstanze.albrecht@uni-bonn.de (K. Albrecht), emma.vonessen@ne.su.sec (E. von Essen), jparys@uni-bonn.de (J. Parys), nora.szech@uni-bamberg.de (N. Szech).

0014-2921/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.euroecorev.2013.02.002

conservative by putting too much weight on the prior,[1] or neglect the prior when they should not.[2] Typical updating problems are mathematically complex, as they require reweighting the prior with the new information. In contrast, in our experiment, Bayes' rule implies that the prior should be ignored completely. Despite the simplicity of the problem, our participants did not follow Bayes' rule. Furthermore, women put more weight on the irrelevant prior than men.[3]

We frame the updating problem in two different ways. In one set of treatments, we used neutral labels ("K" and "L") to distinguish the performer groups. In a second set of treatments, we use gender as the defining property of the two groups: participants evaluate either a man or an equally well performing woman. Evaluators in both setups update more conservatively than rationality would prescribe. Conservatism in updating may consequently be a potential source of the discrimination observed in the labor market, pertaining not only to gender but to any group with less favorable characteristics.

In the economic literature on labor market discrimination, the focus so far has been on two possible rationalizations of discriminatory behavior. Taste-based discrimination (Becker, 1957) assumes that individuals have preferences against interacting with individuals of certain groups, and therefore discriminate against them. In contrast, according to the theory of statistical discrimination (Phelps, 1972; Arrow, 1973), discriminatory behavior arises due to informational frictions. Neither of these theories can explain our findings.[4] The discrimination we observe is non-rational and occurs due to a lack of proper application of basic statistical principles.

The closest previous work to our study was conducted by Reuben et al. (2010), who also try to identify updating problems as a source of discriminatory behavior. They find that in a competitive real effort task, women are less often chosen as leaders than men, even though there are no gender differences in the previous performance. However, although prior performance is known in their experiment, uncertainty about future performance leaves room for statistical discrimination. Furthermore, the chosen leader receives money for being chosen, which may provoke taste-based discrimination. In our study, we eliminate any possibility of such informational frictions and taste-driven behavior; future performance is irrelevant, and no direct or indirect interaction (e.g. via payments) occurs between evaluator and performer. Hence our study neatly shows that conservatism in updating is indeed a source of discriminatory behavior.

As an ex post result, we also find a relationship between self-confidence and gender discrimination: self-confident men tend to overvalue male performers. A potential explanation is that men project their positive attitude towards themselves onto performers of their own gender, but not onto women. In psychology, such overestimation of similarity is well-known under the name of "the false consensus effect" (e.g. Ross et al., 1977; Bauman and Geher, 2002). In general, previous literature indicates that highly self-confident individuals behave differently from individuals with low levels of self-confidence (e.g. Falk et al., 2006; Niederle and Vesterlund, 2007).

We additionally use the results from our study to calibrate a simple model of a job promotion ladder: in each round of this model, employees are promoted with probabilities derived from the evaluations made in the laboratory. We demonstrate a glass ceiling effect, that is, there are virtually no women left after the first few rounds of promotions. In addition, the proportions of women on the different hierarchy levels fit well to observed proportions in the real world (e.g. in academia).

The remainder of the paper is organized as follows: Section 2 introduces the design of our experiment, Section 3 presents our results. Section 4 describes our numerical simulations of the glass ceiling effect, and Section 5 concludes.

2. Design of the experiment

The study consists of two separate stages. In the first stage, participants perform a series of mental rotation tasks (MRTs) and are assigned to one of two groups with different performance averages. The second-stage participants then evaluate the performance of a randomly assigned first-stage subject from one of the two groups. The selection and assignment procedures thereby render prior information on the group performance of one of the two groups irrelevant.[5]

2.1. First stage (pre-study)

The first stage of the study includes 91 participants, called *performers* (50 female, age range: 17 to 49 years, mean age: 23.12 years). Each performer is presented with 24 MRTs. Each MRT consisted of five pictures of three-dimensional objects, one being the original object, and four being rotated or mirrored versions of the original object (adapted from Peters et al., 1995; Vandenberg and Kuse, 1978). The task is completed by indicating which two of the four objects are rotated but not mirrored.

[1] See e.g. Lyon and Slovic, 1976; Bar-Hillel, 1980; Falk et al., 2006; Möbius et al., 2011.

[2] See e.g. Tversky and Kahneman, 1971; Kahneman and Tversky, 1972; Grether, 1992; El-Gamal and Grether, 1995.

[3] Such gender differences in updating are in line with findings from much more demanding updating tasks (Charness and Levin, 2005; Falk et al., 2006; Dohmen et al., 2009; Möbius et al., 2011).

[4] There is a vast empirical literature supporting the existence of discrimination in the labor market. For an overview, see Anderson et al., (2006) and Blau and DeVaro (2007). For example, employers might prefer to rely on group averages rather than bearing the costs of an interview (Anderson et al., 2006).

[5] The study was conducted in the Bonn Econ Lab in Bonn, Germany. Performers were recruited via ORSEE (Greiner, 2003), and mainly consisted of students at Bonn University. Fischbacher's (2007) software zTree was used to present the tasks to the performers.

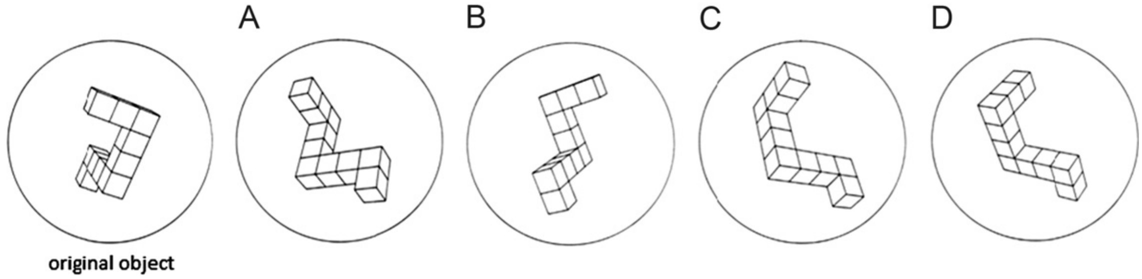


Fig. 1. Example of a mental rotation task presented to the performers. The leftmost object is the original object. Performers have to indicate which two of the four objects (A–D) are rotated but not mirrored versions of the original object. In this example, the correct solutions are B and D.

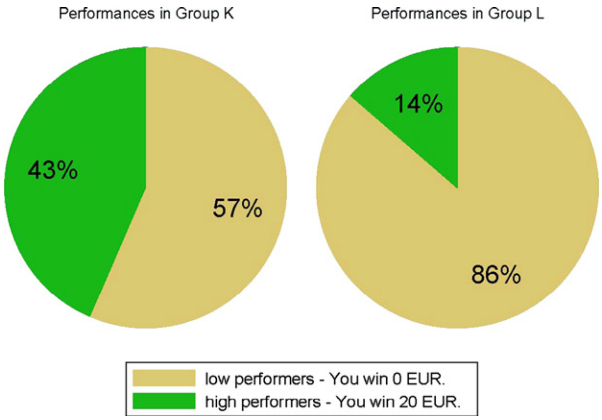


Fig. 2. Pie diagrams presented to the evaluators.

Participants are given two three-minute periods, in each of which they are asked to solve as many as possible of twelve such tasks. Afterwards, we ask the performers to estimate their own performance and the average male and female performance in the task. The performers are assured that all data is treated anonymously. Each performer is paid a flat amount of EUR 2.00 for participation; no performance-based payment was administered on this stage. An example of the task is provided in Fig. 1.

2.2. Second stage (main study)

The second stage of the study includes 356 participants, called *evaluators*. Of these, 351 provide correct answers to all control questions and are therefore included in later analyses (176 women, age range: 19–63 years, mean age: 24.70 years). None of the second-stage participants had previously participated in the first stage. Subjects were recruited from a large field of studies, including arts, economics, law, linguistics, natural sciences, and others.

All second-stage participants are informed about the first stage of the study, and told that they could earn money depending on their own decisions and on the performance of a randomly assigned first-stage participant.

We randomly allocate 308 of the evaluators into four treatments: two neutral treatments, **Neutral** and **Selected-Neutral**, and two gendered treatments, **Man** and **Selected-Woman**. An additional 48 evaluators take part in a control treatment as described below. In total, we ran 15 sessions, each including 20 to 24 subjects.

2.2.1. Neutral treatments

In both neutral treatments, the evaluators are informed via pie diagrams how well the performers from two groups, called group K and group L, performed in the first stage: in group K, 43% were top performers and 57% were mediocre performers, while in group L, 14% were top performers and 86% were mediocre performers (Fig. 2).

The evaluators are further informed that "top performers" solved more than 13 tasks correctly and "mediocre performers" solved at least 9 and at most 13 tasks correctly. They are not asked to perform MRTs themselves, but are presented with the example from Fig. 1.

Table 1
Treatment overview.

Treatment		Description
Neutral treatments	Neutral	Evaluators face a randomly drawn performer from group K
	Selected-neutral	Evaluators face a performer from group L, who is selected as follows: – A performer from group K is randomly chosen – If the performer from group K is a top performer, then a top performer is selected from group L – If the performer from group K is a mediocre performer, then a mediocre performer is selected from group L
Gendered treatments	Man	Evaluators face a randomly drawn male performer
	Selected-woman	Evaluators face a female performer, who is selected as follows: – A male performer is randomly chosen – If the male performer is a top performer, then a female top performer is selected – If the male performer is a mediocre performer, then a female mediocre performer is selected

2.2.2. Selection process in the neutral treatments

The process of assigning first-stage performers is carefully described to the evaluators:

– We randomly draw a first-stage performer from group K.
– We look at the performance level of the drawn performer.
– We draw a performer from group L who performed on the same level.

This means that if the randomly drawn group-K participant is a top performer, then a top performer is selected from group L. If the group-K participant is a mediocre performer, then a mediocre performer is selected from group L.

Evaluators are informed that they will be randomly assigned to a performer from group K (treatment Neutral) or to a selected performer from group L (treatment Selected-Neutral). Thus the probability of facing a top performer is 43% in both treatments. We use control questions to make sure that the evaluators understand the selection procedure (see Appendix A).

2.2.3. Gendered treatments

The gendered treatments are the same as the neutral treatments, except that gender is used as an attribute to define the groups. Hence, here the group-K and group-L performers are labeled male and female performers instead. Accordingly, the treatments are named Man and Selected-Woman. The order of the naming of the groups is counterbalanced throughout the study, i.e. to half of the subjects, group K is presented as the on average better performing group, whereas to the other half of the subjects, group L is presented as the one average better performing group. Thus, we make sure that treatment differences do not occur due to the naming of the groups. An overview of treatments is provided in Table 1.

2.2.4. Evaluation procedure

We elicit the evaluations by letting the evaluators face a series of 50 choices between a certain outcome and a lottery, varying the certain outcome from EUR 0.40 up to EUR 20.00. The lottery outcome depends on the performance of the first-stage performer assigned to the evaluator: If the performer is a top performer, the evaluator wins the lottery (and receives EUR 20.00). If the performer is a mediocre performer, the evaluator loses the lottery (and receives EUR 0.00). The variable of interest is the point of the last risky option before the evaluator switches to the safe option.

2.2.5. Control treatment (Robustness check)

In a control treatment, we address the robustness of our findings from the Selected-Woman treatment. In order to make subjects understand that they may not have to incorporate all statistical information provided to them when working on an evaluation task, we present subjects a sequence of evaluation tasks: In the first task, subjects need statistical information on both female and male performers. In the second task, they only need statistical information on female performers. In the third task, subjects only need statistical information on male performers. The third task is identical to the task in the Selected-Woman treatment. 48 subjects take part in this treatment.

Specifically, in the first task, half of the subjects evaluate a performer who, with 95% probability, is a selected female performer and, with 5% probability, is a randomly drawn female performer. The remaining half of the subjects evaluates a performer who, with 95% probability, is a male and, with 5% probability, is a randomly drawn female performer. Hence, information on both male and female performance is instrumental in this task. In the second task, we ask subjects to estimate the probability that two randomly drawn women are top performers. Thus, in this task only information on female performance is instrumental. The third task is the evaluation task from the Selected-Woman treatment, where only the information on male performance is instrumental.[6,7]

2.2.6. Post-experiment survey

After the evaluators make their choices, they are asked to answer a survey about their socio-demographic background. They are also asked to hypothetically estimate their own performance in the MRTs. At the end of the experiment, all subjects receive a show-up fee of EUR 4.00. Additionally, one of the 50 decisions is randomly drawn for payment. Accordingly, total payments ranged from EUR 4.00 to 24.00 per subject for the gendered and neutral treatments.[8]

3. Results and discussion

We will start by presenting summary statistics of the first stage, showing that there are substantially more male than female top performers in the MRTs. Next, we will analyze the data from the second and main part of our experiment. We will first show that the evaluators are generally conservative in updating, irrationally putting positive weight on the irrelevant prior. Splitting our sample by gender, we will then demonstrate that conservatism is more pronounced among women than among men. The results are essentially the same for the gender-neutral frame and for the gendered frame, with the exception that self-confident men strongly overvalue male performers in the gendered frame.

Performance evaluations are stated in amounts of EUR. All analyses are conducted using two-sided *t*-tests. We also apply the parametric variant of the strategy proposed by Crump et al., (2008) to address potential concerns over multiple testing.[9]

3.1. First stage

There are 43% male top performers, and accordingly 57% male mediocre performers. Only 14% of female performers are top performers, and 86% are mediocre performers. There are significantly more male than female top performers ($\chi^2(1)=4.874$, $p=.027$).

3.2. Second stage

We defined the ‘‘switching point’’ to be the first decision where an evaluator switches from the risky to the safe option; this switching point is considered to be the evaluator’s evaluation of the performer. A total of 33 evaluators switch between the risky and the safe options multiple times, and are therefore excluded from the main analyses.[10] We also exclude three out of 308 evaluators who do not answer all the control questions correctly. Summary statistics of the evaluations by gender and across treatments are provided in Table B1 in Appendix B. In addition, Table B2 lists descriptive statistics by treatment for all but the evaluation variables. Our claim that the randomization into treatments is successful is supported by the fact that Table B2 shows no differences in variables that should not be affected by treatment, such as age, the fraction of participants who switch multiple times, and perceived usefulness of the task.

In both frames, neutral and gendered, we find that evaluators switch significantly earlier in the Selected-Neutral and Selected-Woman treatments compared to the Neutral and Man treatments (neutral treatments: $t(122)=3.766$, $p \ll .001$; gendered treatments: $t(125)=3.239$, $p=.002$); see Fig. 3. This implies that the evaluators are too conservative in updating; that is, they are taking into account the group’s average performance in the Selected-Neutral and Selected-Woman treatments. All evaluators included in the analyses answer all control questions correctly, and hence knew perfectly well that the prior should be irrelevant; but they are not able to update their beliefs accordingly. The fact that this effect is significant in both the neutral and the gendered frame suggests that conservatism in updating is a general phenomenon. That is, conservatism in updating may be a source of discrimination pertaining not only to gender, but to any group with less favorable average characteristics. Table B3 in Appendix B provides summary statistics.

[6] The exact tasks are described in the Instructions in Appendix A. This control treatment was run two years after the other treatments.

[7] Results in the third task are not significantly different depending on the way the first task is formulated (first switching point: $t(41)=0.607$, $p=.547$, average switching point: $t(44)=0.985$, $p=0.330$)

[8] In the control treatment, one of the 50 decisions of the first task as well as one of the 50 decisions of the third task are randomly drawn for payment. Subjects of the control treatment further receive EUR 2.00 if they answer the question in the second task correctly. We pay all tasks of this treatment to ensure that subjects consider all the tasks as equally important.

[9] Two null hypotheses about the average treatment effect are tested. The first hypothesis is that evaluating a performer in the Selected-Neutral treatment instead of a performer in the Neutral treatment has a zero average effect for male and for female evaluators. We test the same hypothesis concerning the gendered treatments. The second hypothesis is that the average treatment effect would be identical for male and female evaluators, i.e. there would be no heterogeneity in the average treatment effect.

[10] As a robustness check, we look at average switching points of all evaluators including multiple switchers. The results are similar for both measures. Although the proportion of female evaluators, who switch multiple times, is significantly higher than among male evaluators ($t(303)=3.154$, $p=.002$), there is no gender difference in the average switching point among evaluators who switch multiple times ($t(31)=1.413$, $p=.168$).

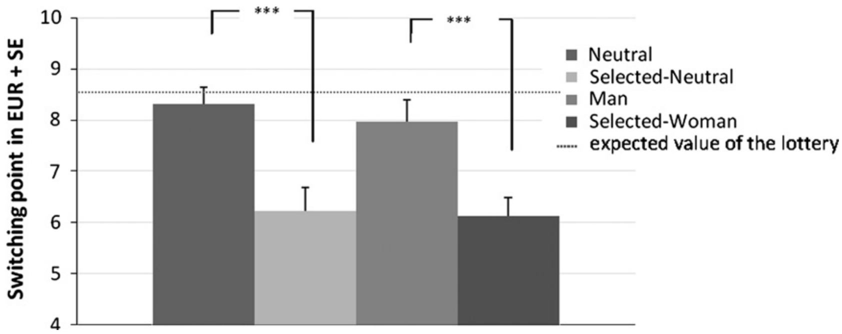


Fig. 3. Evaluations in the four treatments. (SE=Standard Error; ***$p < .01$, **$p < .05$, *$p < .10$, two-sided).

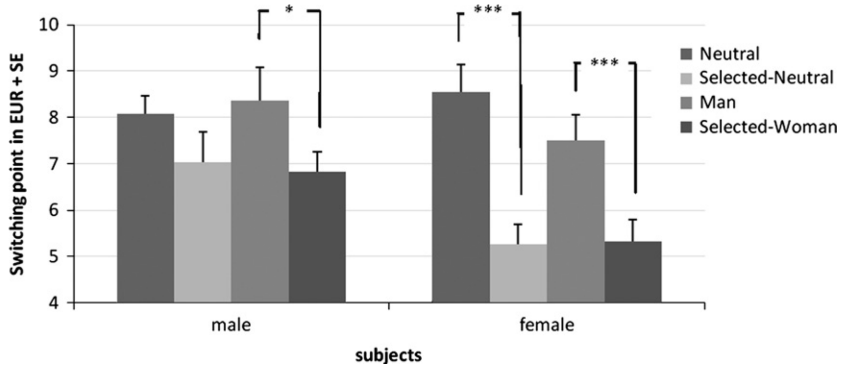


Fig. 4. Evaluations by gender. (SE=Standard Error; ***$p < .01$, **$p < .05$, *$p < .10$, two-sided).

3.2.1. *Gender differences in evaluations between treatments*

Since there is a general gender difference in evaluations ($t(270)=2.441$, $p = .015$) we further analyze them separately for male and female evaluators.

We begin by investigating the general evaluations in the neutral treatments; that is, the evaluations from treatments Neutral and Selected-Neutral. Mean evaluations between these treatments display no significant differences among male evaluators ($t(62)=1.309$, $p=.195$). In contrast, we find a highly significant difference for female evaluators ($t(58)=4.393$, $p < .001$). On average, women evaluate performers from the better group EUR 3.29 higher than they evaluate the selected performers from the worse group. Furthermore, the difference in evaluations between the two neutral treatments is significantly higher for female compared to male evaluators ($t(120)=2.057$, $p=.021$). Hence, women are more conservative in updating than men. An overview of differences in mean evaluations is presented in Fig. 4 and Table B3 in Appendix B.[11]

Like in the neutral treatments, also in the gendered treatments women evaluate the performance of a selected woman significantly lower than the performance of a randomly drawn man ($t(56)=2.772$, $p=.008$). Likewise, for male evaluators we find marginally significant differences in evaluations between the gendered treatments ($t(67)=1.962$, $p=.054$); the difference is marginally significant in the gendered treatments, but not significantly stronger than in the neutral treatments ($t(129) = .450$, $p=.327$). Also female evaluators show no difference in updating between the gendered and the neutral treatments ($t(114)=1.034$, $p=.152$). An overview of differences in mean evaluations is presented in Fig. 4 and Table B3 in Appendix B.[12]

This implies an overall indication of conservatism in updating, and the gendered framing seems to increase this conservatism.

3.2.2. *Control treatment (Robustness check)*

We analyze data from 43 of the 48 subjects who participated in the control treatment. Two subjects were excluded because they did not answer all control questions correctly and three subjects were excluded from the main analysis because they switched multiple times.[13] We compare evaluations from the third task in our control treatment to the Man and Selected-

[11] Tables B4 and B5 in Appendix B provide coefficients estimated from OLS regressions. The results support the findings of the t-tests. In addition, R^2 is.25 in the female regression as compared to only.06 in the male regression, which indicates that differences between treatments explain much more of the variation found in women's evaluations.

[12] In the regressions in Tables B4 and B5, the dummy variable for the gendered treatments and its interactions with a neutral-treatment's dummy and a male dummy are insignificant.

[13] Table B6 in Appendix B provides an overview of the control treatment's results including the data of the three multiple switchers.

Woman treatments of our original experiment to show that our original results are robust even if we make clear before that not all information needs to be incorporated in the evaluation. Evaluations from the control treatment do not differ from evaluations in the Selected-Woman treatment of the original experiment ($\text{mean}_{\text{Control}}=6.24$, $\text{mean}_{\text{Selected-Woman}}=6.13$; $t(126)=0.180$; $p=.858$). In accordance with the data from the Selected-Woman treatment, evaluations differ significantly from those of the Man treatment of the original experiment ($\text{mean}_{\text{Control}}=6.24$, $\text{mean}_{\text{Man}}=7.97$; $t(83)=2.355$; $p=.021$). This again implies an undervaluation of women compared to men.

3.2.3. The influence of self-confidence

In line with previous literature (Falk, Huffman, and Sunde 2006; Niederle and Vesterlund 2007) we find that male evaluators are more optimistic than female evaluators about their own (hypothetical) performance in MRTs ($t(303)=5.781$ $p<.001$). We therefore investigate the influence of self-confidence on discrimination. Our measure of self-confidence is the number of MRTs the evaluators believe they would have solved themselves if they had participated in the first stage. Based on this measure, we construct two groups of evaluators within each treatment and gender: evaluators whose beliefs about their own hypothetical performance are above the median belief (high self-confidence), and those with beliefs below the median (low self-confidence). Fig. 5 and Table B7 in Appendix B give the performance evaluations of male and female evaluators split by level of self-confidence.[14]

The results displayed in Fig. 5 show a clear pattern indicating that self-confident male evaluators discriminate in the gendered treatments ($t(26)=2.079$, $p=.005$), but not in the neutral treatments ($t(24)=.866$, $p=.395$). Importantly, their average evaluation of another man's performance (EUR 10.56) is well above the expected value of the lottery (EUR 8.60). Therefore, self-confident male evaluators in the Man treatment lose money by switching to the safe option too late.

Comparing the mean evaluations between treatments, we conclude that male evaluators with high levels of self-confidence overvalue the performance of other men as opposed to undervaluing the performance of selected women. In contrast, male evaluators with low levels of self-confidence do not overvalue participants from high performing groups (neutral: $t(24)=1.116$, $p=.276$, gendered: $t(26)=0.241$, $p=.812$).[15] We conclude that self-confident men are sensitive to the gender frame, although they know that there are no performance differences.

In line with psychological research on the "false consensus effect", our results may be caused by men projecting their own self-confidence onto other men. The false consensus effect states that individuals overestimate the extent to which others have similar beliefs, opinions, preferences, and habits as they themselves have (e.g. Ross et al., 1977; Bauman and Geher, 2002). Men might consider themselves more similar to other men than to women, and accordingly project their own self-confidence only onto men.

Female evaluators do not display this pattern; that is, self-confidence does not seem to play a role when evaluating the performance of others (Fig. 5).[16] Self-confidence hence seems to be reflected in the behavior of men, but not of women.

4. Practical implications—the glass ceiling

We also apply our results to demonstrate how small differences in male and female evaluation behavior can accumulate to create the glass ceiling phenomenon; that is, the extremely small proportion of female employees seen at higher levels in most professional environments. For this purpose, we consider a simple numerical model of job promotions. This allows us to quantify in a stylistic way how fast the proportion of women decreases when reaching higher levels in professional environments. We also analyze the effects of group size among employees and of proportions of women among the promoters.

4.1. The model

The model includes t hierarchy levels in a firm with n employees at each hierarchy level. Male and female employees exist at each level. Employees at level s are randomly split into m groups of size g. Each group is assigned a male evaluator with probability p_s and a female evaluator otherwise. Men in a group with a male evaluator are assigned a random evaluation drawn from the evaluations of men by male evaluators in the gendered treatments of our experiment. Women

[14] As a robustness check, we alternatively use the mean instead of the median for the sample split. We further create two additional measures of self-confidence. The first of these is a relative self-confidence measure constructed by performing a median-split on the difference between beliefs about one's own performance and the performance of the corresponding performer. The second was constructed by classifying subjects as being self-confident if, according to their beliefs about their own performance, they would count themselves among the group of top performers; that is, according to their belief that they could solve fourteen or more MRTs correctly. Results based on these alternative self-confidence measures are qualitatively similar to those based on the main measure described in the text, and are provided in Tables B8–B10 in Appendix 2.

[15] The estimated coefficients in Tables B11 and B12 confirm this finding. The estimated coefficient of the fourfold interaction dummy for highly self-confident male evaluators in the Man treatment is positive and highly significant. We can also conclude that self-confidence is an important omitted variable in the regression of performance evaluations by male evaluators, as in column 1 of Tables B11 and B12 adding self-confidence improves the fit of the regression from an adjusted R^2 of.03 to one of.19. Also, the coefficient of the dummy variables for being in a non-selected treatment becomes significant only when adding self-confidence.

[16] Accordingly, adding self-confidence only slightly increases the adjusted R^2 from.19 to.23. It does not affect the significance of any estimated coefficient.

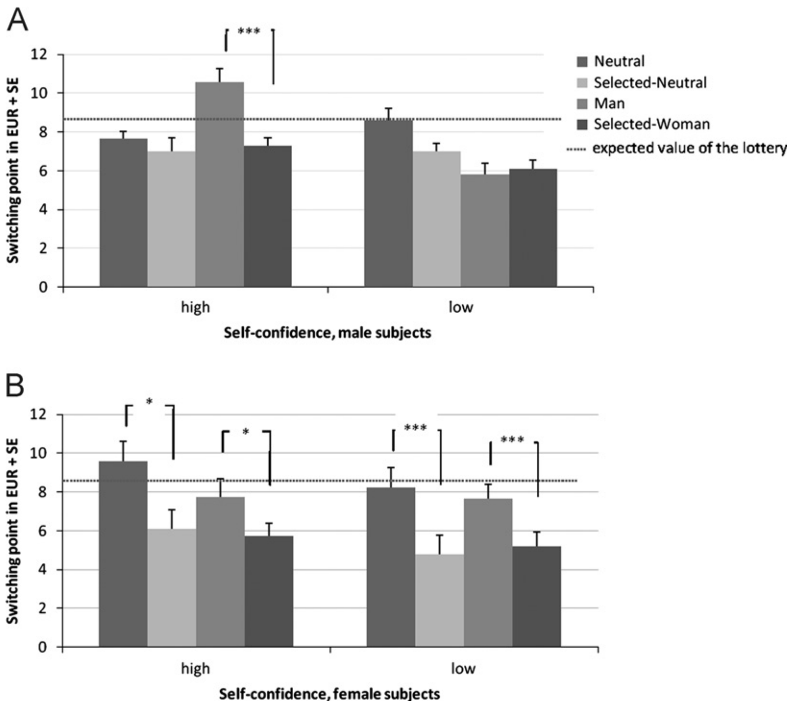


Fig. 5. Evaluations by level of self-confidence of (A) male, and (B) female evaluators. (SE=Standard Error; $^{***}p<.01$, $^{**}p<.05$, $^{*}p<.10$, two-sided).

Table 2
Approximate steady states for 6 hierarchy levels.

0.500	0.416	0.339	0.273	0.216	0.169	g=2
0.500	0.380	0.277	0.194	0.132	0.090	g=4
0.500	0.353	0.230	0.141	0.082	0.045	g=6
0.500	0.332	0.189	0.095	0.045	0.019	g=8

in a group with a male evaluator are assigned a random evaluation drawn from the evaluations of selected women by male evaluators in the gendered treatments of our experiment. Analogously, evaluations in groups with female evaluators are drawn from the evaluations made by female evaluators in the gendered treatments.

In each group, the group member with the highest evaluation is promoted to the next hierarchy level. The number of female employees at level $s+1$ is determined by the number of women promoted at level s. We consider an (approximate) steady state; that is, we choose the proportion p_s of female evaluators at level s to be approximately equal to the proportion of women promoted from level s to level $s+1$. This fixed point is determined by a simple iterative algorithm.[17] We close the model by assuming that the lowest hierarchy level contains equal numbers of female and male employees.

The asymmetry between evaluations of men and selected women in our experiment implies that these promotion dynamics mimic promotions in a professional environment that is traditionally male-dominated. Recall that evaluations are all collected on ex-ante equally-skilled subjects. Thus, we model promotions of equally-skilled and equally-sized male and female populations in a traditionally male-dominated employment field. In order to minimize the contribution of stochastic fluctuations, we average over z runs of these dynamics and choose a number of employees n of sufficient size.[18]

[17] We start with arbitrary proportions of female evaluators (e.g., no female evaluators) and calculate the number of promoted women. These resulting proportions are used as the new proportions of female evaluators. The procedure is iterated until the proportions no longer change significantly. Considering an approximate steady state is justified by the fact that this steady state is usually reached after about three iterations of the procedure.

[18] This implies that the actual size of n is irrelevant as long as n and z are sufficiently large. Notably, we could also consider a promotion pyramid where higher hierarchies are smaller than lower ones. The advantage of equally-sized levels is that computational effort is spread equally across levels.

4.2. Results

Table 2 depicts the approximate steady state proportions of women for different values of g, using 6 hierarchy levels and thus 5 promotions.[19]

As could be expected from our experimental results, we see a moderate decrease in the proportion of women from one hierarchy level to the next. These decreases result in a tiny proportion of women after four or five rounds of promotions. Therefore, the discrimination-driven effects we observed in our experiment are strong enough quantitatively to explain a glass ceiling effect.[20] On comparing our simulations to real-life data on women in academia, we find similar patterns of declining fractions over the hierarchy levels.[21]

5. Conclusion

This study identifies problems in updating as a source of discrimination, beyond taste-based and statistical discrimination. Individuals are not able to fully give up their prior belief concerning groups with known average performances in favor of new individual-specific information. Furthermore, men with high self-confidence about their own performance overestimated the performance of other men.

We also conduct simulations to explore the consequences of our findings in a hierarchical job structure with a sequence of promotion decisions. The results show that the observed discrimination adds up to a glass ceiling effect; that is, to the virtual absence of women after a few rounds of promotions. In line with empirical observations, this effect is even more pronounced when more promotion decisions are made by women.

Our experiment focuses on a benchmark case where the signal about the performance of a person is perfect. For future research it would be interesting to explore conservatism as a possible source of discrimination in other settings, notably, in the field, where signals are less perfect. Separating discrimination due to updating errors from other types of discrimination in the field would likely be a difficult but highly interesting task. Furthermore, it would be of interest to investigate how much better an individual from a less favorable group has to perform in order to receive the same evaluation as another individual with the same abilities.

Acknowledgement

We thank Stefano DellaVigna, Anna Dreber, Armin Falk, Thomas Gall, Magnus Johannesson, Sebastian Kube, Ulrike Malmendier, Benny Moldovanu, Astri Muren, Frank Rosar, Hannah Schildberg-Hörisch, Matthew Rabin, Patrick Schmitz , Nikolaus Schweizer and Matthias Wibral for helpful comments. We further thank Jeroen Jannis Engel, Patrizia Oedyniec, Anne Roesler, Benedikt Vogt, Jana Willrodt, Parwaneh Zekri and Ulf Zoelitz for their assistance. Financial support from the German Research Foundation, via the Bonn Graduate School of Economics (BGSE) the Swedish council for working life and social research (FAS), as well as the Ministry of Innovation, Science, Research, and Technology of NRW is gratefully acknowledged.

Appendix A. Instructions[22]

Sample instructions for gender treatment ("selected-woman"):

Welcome to our study. Please read the following instructions carefully.

For participating in this study you will receive 4 EUR for sure. Depending on you and another participant's performance you can earn money in addition to these 4 EUR. In this study you are anonymous and all data that you provide will be treated confidentially. If you have any questions after reading the following instructions, please raise your hand and we will come to answer your question. Please do not talk to other participants during the study—we would have to exclude you from this study then.

The study consists of two stages: **You are a second stage participant**. During the study, you will be randomly assigned to a participant who participated in the first stage.

[19] The further parameters are $n=2400$ and $z=400$.

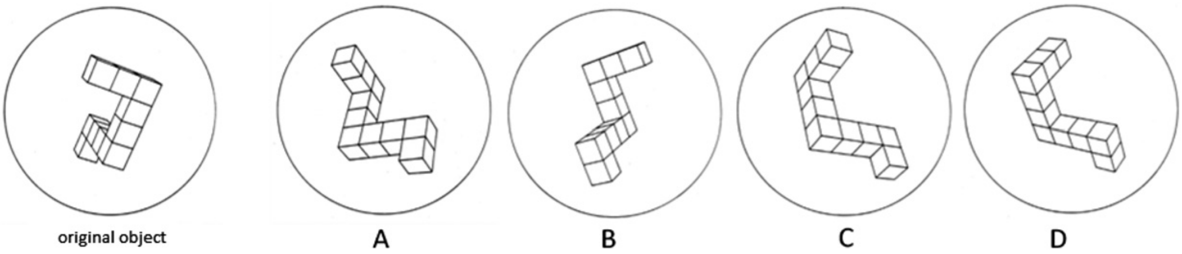
[20] In Appendix C (Table C1), we further demonstrate that the decrease in the number of female employees, when moving up the hierarchy, is more pronounced when promotion decisions are made by women.

[21] For German academia in 2004, the Center of Excellence Women and Science reports the following fractions of women: about 50% among graduates, 39% on the PhD level, 34% among junior researchers, 28% on assistant positions, 14% among full professors, and 9% among the best-paid full professor positions. http://www.bosch-stiftung.de/content/language1/downloads/Kurzexpertise.pdf

[22] The following are an English translation of instructions for the treatments "Selected-Woman" and "Neutral", as well as for the control treatment . The original German versions are available from the authors upon request.

Stage 1

This stage has already been completed by other participants. Those participants solved a number of mental rotation tasks. Here is an example for this task:



In this stage subjects had to distinguish between the two figures among A, B, C, and D that can be transferred into the original object on the left side by rotation (in our example figures B and D). The two other figures (in our example figures A and C) that cannot be transferred into the original object by rotation only, but had to be mirrored. Subjects were supposed to cross out the two figures that were rotations only. If they crossed out both correct figures, the task was solved correctly. Subjects were given 24 of these tasks, and 6 minutes to solve as many of them as possible.

Stage 2

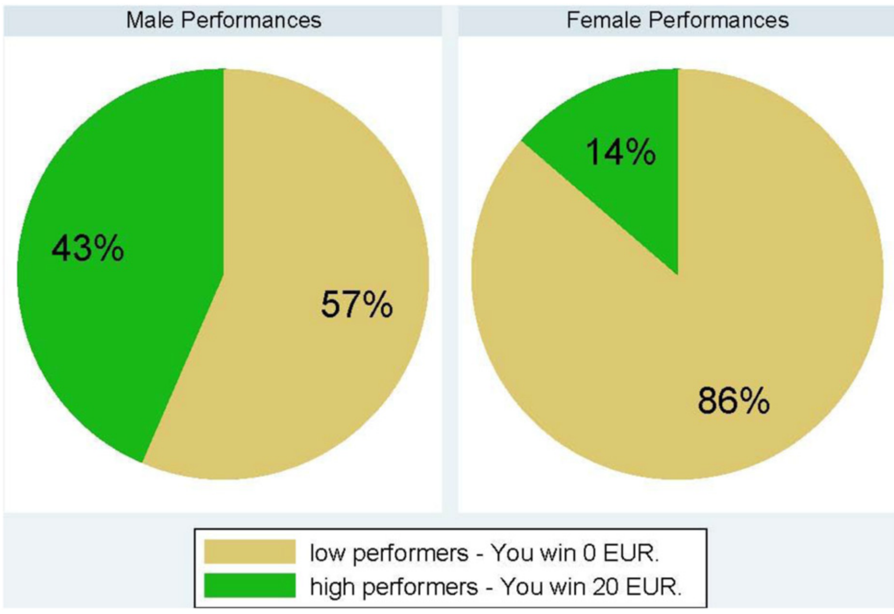
In this stage **you** are the sole decision maker. You have the opportunity to earn money depending on the stage 1 performance of a first stage participant you have been randomly matched with. The payment is for you only, the first stage participant was paid for his participation already.

In the following, we only regarded participants of the first stage who solved a minimum number of tasks correctly. participants were divided into two groups, males and females.

– 43% of the male participants are top perfomers. 57% are mediocre performers.
– 14% of the female participants are top perfomers. 86% are mediocre performers.

Top performers are participants who solved 14 or more tasks correctly, whereas mediocre performers are participants who solved at least 9 tasks correctly, but not more than 13 tasks.

Below we present the distributions of the two groups in the form of a diagram.



Please answer the following control questions:

1. Imagine that there were 100 male participants in Stage 1. What is the number of male top performers?_____ participants are top performers.
2. Imagine that there were 100 female participants in Stage 1. What is the number of female top performers?_____ participants are top performers.
3. Imagine that there were 100 male participants in Stage 1. What is the number of male mediocre performers?_____ participants are mediocre performers.
4. Imagine that there were 100 female participants in Stage 1. What is the number of female mediocre performers?_____ participants are mediocre performers.

The selection of the female and male participants:

Please read the following paragraph carefully. It is important that you understand the selection process.

The **selection** of the first stage participants was as follows:

We randomly select one male participant. We call him participant M henceforth. Then we will select a female participant F as follows:

– If participant M was a top performer, we select a female participant F who also was a top performer.
– If participant M was a mediocre performer, we choose a female participant F who also was a mediocre performer.

You will get matched either to the male particioant M or the female participant F.

Later, you will choose between a fixed reward and a lottery:

– You receive 20 EUR if the participant you are matched with is a top performer.
– You receive 0 EURif the participant is a mediocre performer.

Please answer the following control questions:

Imagine that one male and one female participant are selected as it is described above. Please indicate by putting an X which alternative you think is correct in the following two situations.

1. If the male participant is a top performer, then the female participant is a
   —— top performer —— mediocre performer —— could be either or
2. If the male participant is a mediocre performer, then the female participant is a
   ___ top performer ____ mediocre performer ___ could be either or

Please insert the correct answer in the following two situations.

1. If the person you got matched to is a top performer, you receive _____ EUR.
2. If the person you got matched to is a mediocre performer, you receive _____ EUR.

Decision

We now ask you to make a decision for each of the following options between getting a fixed amount of money(from 0.40 EUR going up to 20 EUR), and playing the aforementioned lottery.

At the end, one of your decisions will be randomly drawn and determine your final payoff.

Mark your answers by putting an X at the alternative you choose for each of the questions 1 to 50.

1. When a decision will be drawn in which you chose the fixed reward, you will receive this reward.
2. When a decision will be drawn in which you chose the lottery, you will receive 0 or 20 EUR, depending on the performance of your matched first stage participant.
3. If you do not put an X in the decision that was drawn, you will receive 0 EUR.

According to selection process described above, you have been matched with a female participant.

Making these decisions we ask you to take your time to think about your decisions and to take them seriously. **Also, remember that you will be paid according to one of these decisions, which is randomly drawn after the study ends.**

(1) Which alternative do you choose:	
☐ 0.40 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
(2) Which alternative do you choose:	
☐ 0.80 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
(3) Which alternative do you choose:	
☐ 1.20 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
….	
(50) Which alternative do you choose:	
☐ 20.00 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant

Sample instructions for neutral treatment ("neutral"):

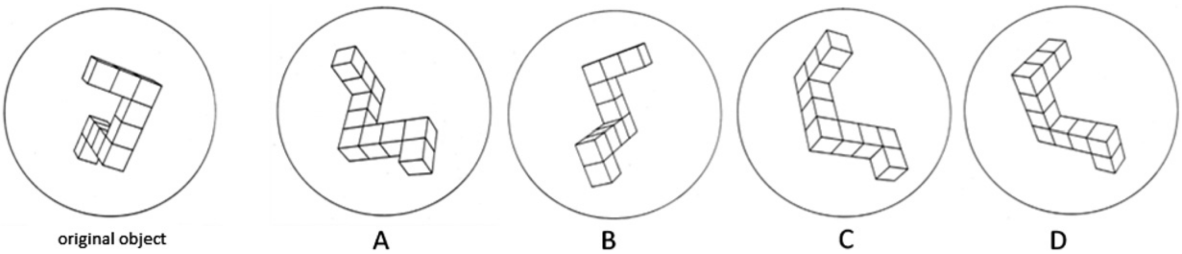
Welcome to our study. Please read the following instructions carefully.

For participating in this study you will receive 4 EUR for sure. Depending on you and another participant's performance you can earn money in addition to these 4 EUR. In this study you are anonymous and all data that you provide will be treated confidentially. If you have any questions after reading the following instructions, please raise your hand and we will come to answer your question. Please do not talk to other participants during the study – we would have to exclude you from this study then.

The study consists of two stages: **You are a second stage participant**. During the study, you will be randomly assigned to a participant who participated in the first stage.

Stage 1

This stage has already been completed by other participants. Those participants solved a number of mental rotation tasks. Here is an example for this task:



In this stage subjects had to distinguish between the two figures among A, B, C, and D that can be transferred into the original object on the left side by rotation (in our example figures B and D). The two other figures (in our example figures A and C) that cannot be transferred into the original object by rotation only, but had to be mirrored. Subjects were supposed to cross out the two figures that were rotations only. If they crossed out both correct figures, the task was solved correctly. Subjects were given 24 of these tasks, and 6 min to solve as many of them as possible.

Stage 2

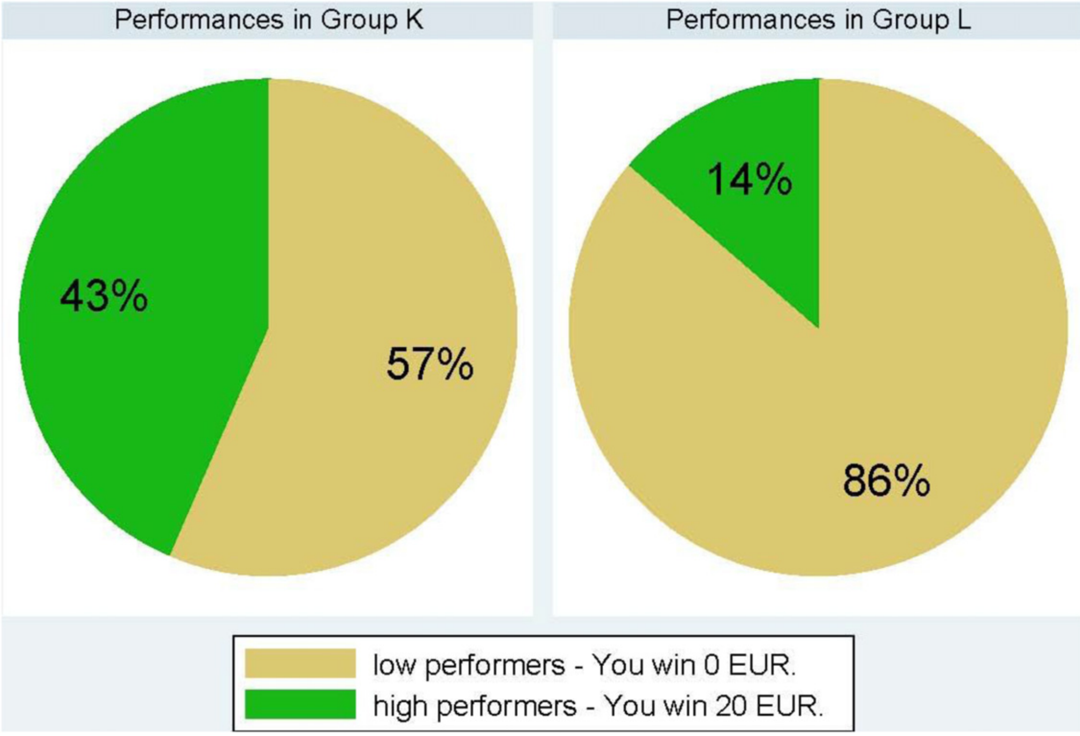
In this stage **<u>you</u>** are the sole decision maker. You have the opportunity to earn money depending on the stage 1 performance of a first stage participant you have been randomly matched with. The payment is for you only, the first stage participant was paid for his participation already.

In the following, we only regarded participants of the first stage who solved a minimum number of tasks correctly. participants were divided into two groups, K and L.

- 43% participants of group K are top perfomers. 57% are mediocre performers.
- 14% participants of group L are top perfomers. 86% are mediocre performers.

Top performers are participants who solved 14 or more tasks correctly, whereas mediocre performers are participants who solved at least 9 tasks correctly, but not more than 13 tasks.

Below we present the distributions of the two groups in the form of a diagram.



Please answer the following control questions:

1. Imagine that there were 100 participants in group K in Stage 1. What is the number of group K top performers?_____ participants are top performers.
2. Imagine that there were 100 participants in group L in Stage 1. What is the number of group L top performers?_____ participants are top performers.
3. Imagine that there were 100 participants in group K in Stage 1. What is the number of group K mediocre performers?_____ participants are mediocre performers.
4. Imagine that there were 100 participants in group L in Stage 1. What is the number of group L mediocre performers?_____ participants are mediocre performers.

The selection of the participants from group K and L:

Please read the following paragraph carefully. It is important that you understand the selection process.

The **selection** of the first stage participants was as follows:

We randomly select one participant from group K. We call him participant k henceforth. Then we will select a participant L as follows:

– If participant K was a top performer, we select a participant L who also was a top performer.
– If participant K was a mediocre performer, we choose a participant L who also was a mediocre performer.

You will get matched either to the male participant K or participant L.

Later, you will choose between a fixed reward and a lottery:

– You receive 20 EUR if the participant you are matched with is a top performer.
– You receive 0 EUR if the participant is a mediocre performer.

Please answer the following control questions:

Imagine that one participant K and one participant *L* are selected as it is described above. Please indicate by putting an X which alternative you think is correct in the following two situations.

1. If participant K is a top performer, then participant L is a
   ___ top performer ____ mediocre performer ___ could be either or
2. If participant K is a mediocre performer, then participant L is a
   ___ top performer ____ mediocre performer ___ could be either or

Please insert the correct answer in the following two situations.

1. If the person you got matched to is a top performer, you receive _____ EUR.
2. If the person you got matched to is a mediocre performer, you receive _____ EUR.

Decision

We now ask you to make a decision for each of the following options between getting a fixed amount of money (from 0.40 EUR going up to 20 EUR), and playing the aforementioned lottery.

At the end, one of your decisions will be randomly drawn and determine your final payoff.

Mark your answers by putting an X at the alternative you choose for each of the questions 1 to 50.

1. When a decision will be drawn in which you chose the fixed reward, you will receive this reward.
2. When a decision will be drawn in which you chose the lottery, you will receive 0 or 20 EUR, depending on the performance of your matched first stage participant.
3. If you do not put an X in the decision that was drawn, you will receive 0 EUR.

According to selection process described above, you have been matched with a participant from group K.

Making these decisions we ask you to take your time to think about your decisions and to take them seriously. **Also, remember that you will be paid according to one of these decisions, which is randomly drawn after the study ends.**

(1) Which alternative do you choose:	
☐ 0.40 EUR for sure	☐ 0 or 20 EUR depending on the performance of your participant
(2) Which alternative do you choose:	
☐ 0.80 EUR for sure	☐ 0 or 20 EUR depending on the performance of your participant
(3) Which alternative do you choose:	
☐ 1.20 EUR for sure	☐ 0 or 20 EUR depending on the performance of your participant
….	
(50) Which alternative do you choose:	
☐ 20.00 EUR for sure	☐ 0 or 20 EUR depending on the performance of your participant

Sample instructions for control treatment:

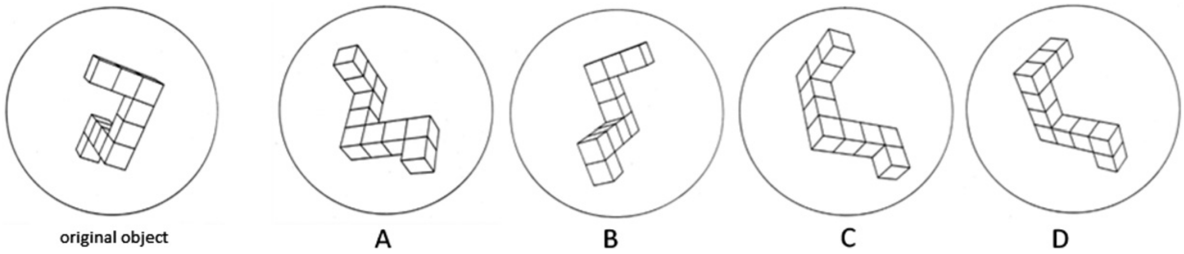
Welcome to our study. Please read the following instructions carefully.

For participating in this study you will receive 4 EUR for sure. Depending on you and another participant's performance you can earn money in addition to these 4 EUR. In this study you are anonymous and all data that you provide will be treated confidentially. If you have any questions after reading the following instructions, please raise your hand and we will come to answer your question. Please do not talk to other participants during the study – we would have to exclude you from this study then.

The study consists of two stages: **You are a second stage participant**. During the study, you will be randomly assigned to a participant who participated in the first stage.

Stage 1

This stage has already been completed by other participants. Those participants solved a number of mental rotation tasks. Here is an example for this task:



In this stage subjects had to distinguish between the two figures among A, B, C, and D that can be transferred into the original object on the left side by rotation (in our example figures B and D). The two other figures (in our example figures A and C) that cannot be transferred into the original object by rotation only, but had to be mirrored. Subjects were supposed to cross out the two figures that were rotations only. If they crossed out both correct figures, the task was solved correctly. Subjects were given 24 of these tasks, and 6 min to solve as many of them as possible.

Stage 2

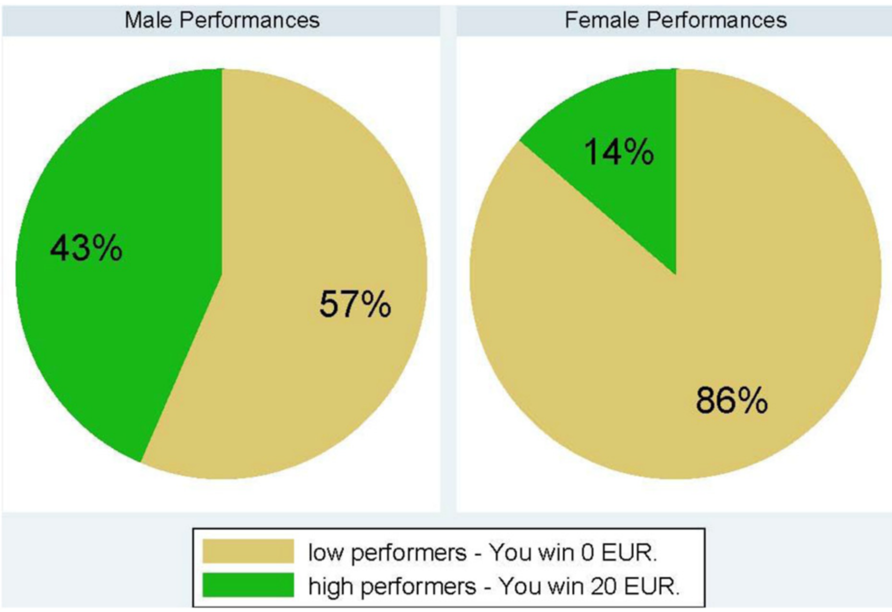
In this stage **you** are the sole decision maker. You have the opportunity to earn money depending on the stage 1 performance of a first stage participant you have been randomly matched with. The payment is for you only, the first stage participant was paid for his participation already.

In the following, we only regarded participants of the first stage who solved a minimum number of tasks correctly. participants were divided into two groups, males and females.

- 43% of the male participants are top perfomers. 57% are mediocre performers.
- 14% of the female participants are top perfomers. 86% are mediocre performers.

Top performers are participants who solved 14 or more tasks correctly, whereas mediocre performers are participants who solved at least 9 tasks correctly, but not more than 13 tasks.

Below we present the distributions of the two groups in the form of a diagram.



Please answer the following control questions:

1. Imagine that there were 100 male participants in Stage 1. What is the number of male top performers?_____ participants are top performers.
2. Imagine that there were 100 female participants in Stage 1. What is the number of female top performers?_____ participants are top performers.
3. Imagine that there were 100 male participants in Stage 1. What is the number of male mediocre performers?_____ participants are mediocre performers.
4. Imagine that there were 100 female participants in Stage 1. What is the number of female mediocre performers?_____ participants are mediocre performers.

The selection of the female and male participants:

Please read the following paragraph carefully. It is important that you understand the selection process.

The **selection** of the first stage participants was as follows:

We randomly select one male participant. We call him participant M henceforth. Then we will select a female participant F as follows:

– If participant M was a top performer, we select a female participant F who also was a top performer.
– If participant M was a mediocre performer, we choose a female participant F who also was a mediocre performer.

You will get matched either to the male participant M or the female participant F with a probability of 95%. Otherwise (with a probability of 5%), you will get matched to a randomly drawn female participant. Before you will make your decision, you will learn whether you were matched to participant M or F with a probability of 95%.

Later, you will choose between a fixed reward and a lottery:

– You receive 20 EUR if the participant you are matched with is a top performer.
– You receive 0 EUR if the participant is a mediocre performer.

Please answer the following control questions:

Imagine that one male and one female participant are selected as it is described above. Please indicate by putting an X which alternative you think is correct in the following two situations.

1. If the male participant participant is a top performer, then the female participant is a
   ___ top performer ____ mediocre performer ___ could be either or
2. If the male participant is a mediocre performer, then the female participant is a
   ___ top performer ____ mediocre performer ___ could be either or

Please insert the correct answer in the following two situations.

1. If the person you got matched to is a top performer, you receive _____ EUR.
2. If the person you got matched to is a mediocre performer, you receive _____ EUR.

Decision

We now ask you to make a decision for each of the following options between getting a fixed amount of money (from 0.40 EUR going up to 20 EUR), and playing the aforementioned lottery.

At the end, one of your decisions will be randomly drawn and determine your final payoff.

Mark your answers by putting an X at the alternative you choose for each of the questions 1 to 50.

1. When a decision will be drawn in which you chose the fixed reward, you will receive this reward.
2. When a decision will be drawn in which you chose the lottery, you will receive 0 or 20 EUR, depending on the performance of your matched first stage participant.
3. If you do not put an X in the decision that was drawn, you will receive 0 EUR.

According to selection process described above, you have been matched with a participant. With a probability of 95%, this participant is male participant M. With a probability of 5%, this participant is a randomly drawn female participant.

Making these decisions we ask you to take your time to think about your decisions and to take them seriously. **Also, remember that you will be paid according to one of these decisions, which is randomly drawn after the study ends.**

(1) Which alternative do you choose:	
☐ 0.40 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
(2) Which alternative do you choose:	
0.80 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
(3) Which alternative do you choose:	
☐ 1.20 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
….	
(50) Which alternative do you choose:	
☐ 20.00 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant

Please answer the following question:

The computer randomly draws two female participants. How high is the probability that two top performing female participants will be drawn? If you answer correctly, you will receive € 2.

a) Below 25%
b) Above 25%, and below 50%
c) Above 50%, and below 75%
d) Above 75%

The selection of the female and male participants:

Please read the following paragraph carefully. It is important that you understand the selection process. Please note that this selection process differs from the one described before!

The **selection** of the first stage participants was as follows:

We randomly select one male participant. We call him participant M henceforth. Then we will select a female participant F as follows:

- If participant M was a top performer, we select a female participant F who also was a top performer.
- If participant M was a mediocre performer, we choose a female participant F who also was a mediocre performer.

This time, you will get matched either to the male participant *M* or the female participant F with a probability of 100%.

Later, you will choose between a fixed reward and a lottery:

- You receive 20 EUR if the participant you are matched with is a top performer
- You receive 0 EUR if the participant is a mediocre performer

Please answer the following control questions:

Imagine that one male and one female participant are selected as it is described above. Please indicate by putting an X which alternative you think is correct in the following two situations.

1. If the male participant participant is a top performer, then the female participant is a
   ___ top performer ____ mediocre performer ___ could be either or
2. If the male participant is a mediocre performer, then the female participant is a
   ___ top performer____ mediocre performer ___ could be either or

Please insert the correct answer in the following two situations.

1. If the person you got matched to is a top performer, you receive _____ EUR.
2. If the person you got matched to is a mediocre performer, you receive _____ EUR.

Decision

We now ask you to make a decision for each of the following options between getting a fixed amount of money (from 0.40 EUR going up to 20 EUR), and playing the aforementioned lottery.

At the end, one of your decisions will be randomly drawn and determine your final payoff.

Mark your answers by putting an X at the alternative you choose for each of the questions 1 to 50.

1. When a decision will be drawn in which you chose the fixed reward, you will receive this reward.
2. When a decision will be drawn in which you chose the lottery, you will receive 0 or 20 EUR, depending on the performance of your matched first stage participant.
3. If you do not put an X in the decision that was drawn, you will receive 0 EUR.

According to selection process described above, you have been matched with a female participant.

Making these decisions we ask you to take your time to think about your decisions and to take them seriously. **Also, remember that you will be paid according to one of these decisions, which is randomly drawn after the study ends.**

(1) Which alternative do you choose:	
☐ 0.40 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
(2) Which alternative do you choose:	
☐ 0.80 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
(3) Which alternative do you choose:	
☐ 1.20 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant
….	
(50) Which alternative do you choose:	
☐ 20.00 EUR for sure	☐ 0 or 20 EUR depending on the performance of your female participant

Appendix B

See Tables B1–B12 here.

Table B1
Descriptive statistics for evaluators by gender.

Subjects variable	All		Male		Female		Gender
	Mean (Std. dev.)	Median Obs.	Mean (Std. dev.)	Mesdian Obs.	Mean (Std. dev.)	Median Obs.	Difference in means
First switching point[a] in EUR	6.84	6.8	7.28	7.60	6.35	5.60	0.93**
	(3.18)	272	(3.14)	144	(3.16)	128	[0.015]
Av. switching point in EUR	6.98	6.8	7.39	7.60	6.57	6.00	0.82**
	(3.18)	305	(3.17)	152	(3.14)	153	[0.023]
Multiple switcher (dummy)	0.11	0	0.05	0	0.16	0	−0.11***
	(0.31)	305	(0.22)	152	(0.37)	153	[0.002]
0.5							
Belief: own score	14.00	14.0	15.36	16	12.64	12	2.72***
	(4.32)	305	(4.17)	152	(4.05)	153	[0.000]
Belief: participant's score	13.17	14.0	13.36	14	12.99	14	0.37
	(3.11)	305	(3.25)	152	(2.97)	153	[0.302]
Diff. in beliefs: own participant	0.83	0.0	2.01	2	−0.35	0	2.36***
	(4.24)	305	(3.80)	152	(4.34)	153	[0.000]
Belief: av.male score	14.16	14.0	13.92	13	14.41	14	−0.49
	(2.92)	305	(2.99)	152	(2.83)	153	[0.148]
Belief: av.female score	10.87	10.0	10.97	11	10.77	10	0.20
	(2.87)	305	(2.99)	152	(2.74)	153	[0.552]
Task usefulness (1=low to 10=high)	6.52	7.0	6.64	7	6.41	7	0.23
	(2.33)	305	(2.35)	152	(2.31)	153	[0.371]
Age	24.70	24.0	25.36	25	24.04	24	1.32**
	(4.59)	305	(5.27)	152	(3.69)	153	[0.012]

***, **, *: Difference is significant on the 1, 5, 10 percent level (two-sided t test).
[a] Subjects with more than one switching point are excluded in all tables.

Table B2
Descriptive statistics for evaluators by treatment.

Treatment **Variable**	**Man** 45 Obs.	**Selected woman** 94 Obs	**Neutral** 70 Obs	**Selected neutral** 69 Obs.	**Difference**[a] *p*-value
Multiple switcher as dummy	0.13	0.10	0.11	0.10	0.95
	(0.33)	(0.30)	(0.32)	(0.30)	
Belief: own score	13.35	14.12	14.17	14.29	0.68
	(4.16)	(4.76)	(3.45)	(4.86)	
Belief: participant's score	13.79	12.24	14.39	13.54	0.00
	(3.24)	(2.99)	(2.46)	(3.31)	
Diff. in beliefs: own participant	−0.44	1.87	−0.21	0.75	0.00
	(4.64)	(4.65)	(3.22)	(4.05)	
Belief: av.male score	13.83	14.56	14.04	14.06	0.46
	(2.59)	(2.71)	(2.85)	(3.45)	
Belief: av.female score	9.67	10.44	11.63	12.03	0.00
	(2.38)	(2.67)	(2.61)	(3.23)	
Task usefulness	6.42	6.51	6.67	6.77	0.83
(1=low to 10=high)	(2.47)	(2.41)	(2.21)	(2.26)	
Age	24.98	24.40	24.24	24.97	0.72
	(3.26)	(4.97)	(3.17)	(6.06)	

Reported are means and standard deviations in parentheses. Switching points (evaluations) are reported separately in Table B3.

[a] *P*-value of Wald test on joint significance of β, γ, and δ in linear regressions of the form: Variable$=\alpha+\beta$*(Treament Selected woman)$+\gamma$*(Treament Neutral)$+\delta$*(Treament Selected neutral)$+\varepsilon$.

Table B3
Performance evaluations.

Treatment		First switching point in EUR			Average switching point in EUR		
	Subjects	**All**	**Male**	**Female**	**All**	**Male**	**Female**
Man	Mean (Std.err.)	7.97 (0.46)	8.37 (0.70)	7.49 (0.57)	7.86 (0.41)	8.30 (0.67)	7.42 (0.49)
	Median	7.6	8.0	6.8	7.6	8.0	7.2
	Obs.	42	23	19	48	24	24
Selected-woman	Mean (Std.err.)	6.13 (0.33)	6.82 (0.43)	5.32 (0.47)	6.32 (0.32)	7.11 (0.46)	5.47 (0.43)
	Median	5.6	7.2	4.8	5.6	7.6	4.8
	Obs.	85	46	39	94	49	45
Difference in means[a]		1.84***[.001]	1.55*[.054]	2.17***[.008]	1.53***[.005]	1.19 [.143]	1.95***[.006]
Zero ATE[b]		$\chi^2(2)=12.163$ [.002]***			$\chi^2(2)=11.160$ [.004]***		
Constant ATE[b]		$\chi^2(1)=0.318$ [.573]			$\chi^2(1)=0.536$ [.464]		
Neutral	Mean (Std.err.)	8.32 (0.36)	8.07 (0.39)	8.55 (0.59)	8.29 (0.34)	8.04 (0.40)	8.51 (0.53)
	Median	8.2	7.8	8.4	8.2	7.8	8.4
	Obs.	62	30	32	70	32	38
Selected-neutral	Mean (Std.err.)	6.23 (0.43)	7.02 (0.66)	5.26 (0.44)	6.51 (0.41)	7.07 (0.64)	5.93 (0.49)
	Median	5.6	6.8	5.0	6.0	6.8	5.6
	Obs.	62	34	28	69	35	34
Difference in means[a]		2.09***[.000]	1.05 [.195]	3.29***[.000]	1.78***[.001]	0.97 [.218]	2.58***[.001]
Zero ATE [b]		$\chi^2(2)=22.071$ [.000]***			$\chi^2(2)=14.455$ [.001]***		
Constant ATE[b]		$\chi^2(1)=4.476$ [.034]***			$\chi^2(1)=2.386$ [.123]		

***, **, *: Difference is significant on the 1, 5, 10 percent level. Standard error in parentheses, *p*-values in brackets.

[a] Two-sided *t* test.

[b] Tests for treatment effect heterogeneity as in Crump, Hotz, Imbens, and Mitnik (2008). The first (second) is testing whether facing a selected person has a zero (an identical) average effect for male and female subjects.

Table B4
Regression analysis of performance evaluations by gender.

Subjects	Switching point		
	Male	**Female**	**All**
Non-select TM * gender TM * male			1.61
			(1.52)
Non-select TM * male			−2.24**
			(1.02)
Gender TM * male			−.26
			(.99)
Non-select TM * gender TM	.33	−1.04	−1.12
	(1.10)	(1.03)	(1.07)
Male			1.75**
			(.74)
Non-select TM	1.16*	3.24***	3.30***
	(.71)	(.72)	(.71)
Gender TM	−.11	.05	.07
	(.76)	(.61)	(.65)
Age	.07	−.16**	.007
	(.12)	(.07)	(.09)
Constant	5.13*	9.02**	5.09**
	(2.97)	(1.73)	(2.26)
Obs.	133	118	251
R squared	.06	.25	.14

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Non-select TM is a dummy variable equal to one for the treatments *Man* and *Neutral*, and zero otherwise. Gender TM is a dummy that equals one for the gendered treatments *Man* and *Selected-woman*.
***, **, *: Difference is significant on the 1, 5, 10 percent level. Standard errors in parentheses.

Table B5
Regression analysis of performance evaluations by gender.

Subjects:	Average switching point		
	Male	**Female**	**All**
Non-select TM * gender TM * male			.86(1.47)
Non-select TM * male			−1.62(1.00)
Gender TM * male			.49(.99)
Non-select TM * gender TM	.05(1.03)	−.64(.92)	−.63(.95)
Male			1.15(.71)
Non-select TM	1.07(.67)	2.61***(.68)	2.58***(.70)
Gender TM	.14(.73)	−.40(.61)	−.46(.66)
Age	.07(.12)	−.16***(.06)	−.004(.09)
Constant	5.33*(2.96)	9.80***(1.54)	6.02***(2.17)
Obs.	140	141	281
R squared	.04	.20	.11

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Non-select TM is a dummy variable equal to one for the treatments *Man* and *Neutral*, and zero otherwise. Gender TM is a dummy that equals one for the gendered treatments *Man* and *Selected-Woman*.
***, **, *: Difference is significant on the 1, 5, 10 percent level. Standard errors in parentheses.

Table B6 presents an overview of all tasks used in the control treatment. The third task is identical to the task subjects face in the Selected-woman treatment. The other two tasks were chosen to ensure that subjects understand that not in every evaluation task, all statistical information has to be incorporated.

Table B6
Performance evaluations in the control treatment.

Treatment	**Subjects**	First switching point in EUR: **All**	**Male**	**Female**	Average switching point in EUR: **All**	**Male**	**Female**
Third Task (as in selected-woman treatment)	Mean (Std.err.)	6.24 (0.57)	6.13 (0.72)	6.35 (0.89)	6.37 (0.56)	6.13 (0.72)	6.56 (0.86)
	Median	5.6	5.6	5.2	5.6	5.6	5.2
	Obs.	43	21	22	46	21	25

		First switching point in EUR						Average switching point in EUR					
		All		**Male**		**Female**		**All**		**Male**		**Female**	
		1A[a]	**1B**	**1A**	**1B**	**1A**	**1B**	**1A**	**1B**	**1A**	**1B**	**1A**	**1B**
First task	Mean (Std.err.)	7.55 (0.81)	5.85 (0.99)	6.84 (0.92)	5.68 (1.46)	8.60 (1.49)	5.95 (1.40)	7.77 (0.80)	6.56 (0.90)	6.40 (0.94)	8.00[b] (1.57)	9.29 (1.18)	5.60 (1.04)
	Median	7.6	4.8	5.6	4.0	8.0	5.2	7.6	6.2	5.6	8.6	8.4	5.2
	Obs.	15	13	9	5	6	8	19	20	10	8	9	12

		Fraction of correct solutions: **All**	**Male**	**Female**
Second task	Mean (Std.err.)	.83 (.05)	.91 (.06)	.77 (.08)
	Obs.	48	24	24

[a] In task 1A, with 95% probability a male performer, and with 5% probability a randomly drawn female performer is selected. In task 1B, with 95% probability a selected female performer, and with 5% probability a randomly drawn female performer is selected.

[b] Note that out of the eight male evaluators here, two (who both were multiple switchers) have extremely late average switching points, at EUR 12.00 and at EUR 15.20, respectively (the risk-neutral value of the lottery is EUR 8.31). Leaving out these two extreme data points, the mean is EUR 6.13.

Table B7
Performance evaluations by relative level of self-confidence.

		First switching point in EUR				Average switching point in EUR			
		Male subjects		**Female subjects**		**Male subjects**		**Female subjects**	
	Self-confi-dence[c]	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Man	Mean (Std.err.)	10.56 (0.94)	5.80 (0.68)	7.75 (0.70)	7.65 (1.17)	10.56 (0.94)	5.91 (0.61)	7.38 (0.72)	7.71 (0.88)
	Median	9.6	5.6	8.4	6.8	9.6	5.6	8.0	6.8
	Obs.	10	8	8	8	10	9	9	11
Selected-woman	Mean (Std.err.)	7.27 (0.60)	6.08 (0.68)	5.72 (0.78)	5.18 (0.76)	7.27 (0.60)	6.38 (0.65)	6.13 (0.73)	5.18 (0.65)
	Median	7.8	5.6	4.8	4.4	7.8	6.4	5.6	4.2
	Obs.	18	20	13	18	18	22	15	22
Difference in means[a]		3.29***	−0.28	2.03*	2.47*	3.29***	−0.47	1.25	2.53**
[p-value]		[.005]	[.812]	[.090]	[.085]	[.005]	[.671]	[.271]	[.029]
Zero ATE[b]		$\chi^2(2)=8.692$[.013]**		$\chi^2(2)=6.949$[.031]**		$\chi^2(2)=8.931$ [.012]**		$\chi^2(2)=6.826$ [.033]**	
Constant ATE[b]		$\chi^2(1)=5.849$[.016]**		$\chi^2(1)=0.066$[.798]		$\chi^2(1)=6.901$ [.009]***		$\chi^2(1)=0.735$ [.391]	
Neutral	Mean (Std.err.)	7.65 (0.55)	8.62 (0.63)	9.6 (0.94)	8.24 (0.73)	7.65 (0.55)	8.80 (0.60)	9.46 (0.82)	8.31 (0.70)
	Median	7.6	8.4	9.6	8.4	7.6	8.4	9.6	8.4
	Obs.	15	11	12	15	15	12	14	18
Selected-Neutral	Mean (Std.err.)	7.02 (0.42)	6.99 (1.16)	6.08 (0.65)	4.77 (0.76)	7.02 (0.42)	7.10 (1.09)	6.62 (0.75)	5.34 (0.79)
	Median	7.6	7.6	6.4	4.2	7.6	7.6	6.8	4.6
	Obs.	11	15	10	12	11	16	13	14
Difference in means[a]		0.63	1.63	3.52***	3.47***	0.63	1.70	2.84**	2.97***
[p-value]		[.395]	[.276]	[.008]	[.003]	[.395]	[.225]	[.017]	[.009]
Zero ATE[b]		$\chi^2(2)=2.375$ [.305]		$\chi^2(2)=20.303$[.000]***		$\chi^2(2)=2.704$ [.259]		$\chi^2(2)=14.484$ [.001]***	
Constant ATE[b]		$\chi^2(1)=0.447$ [.504]		$\chi^2(1)=0.001$[.976]		$\chi^2(1)=0.557$ [.455]		$\chi^2(1)=0.007$ [.934]	

***, **, *: Difference is significant on the 1, 5, 10 percent level. p-values in brackets.
[a] Two-sided t test.
[b] Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical) average effect for highly and less self-confident subjects.
[c] Self-confidence is high (low) if the belief about own score is above (below) the median belief by gender.

Table B8
Performance evaluations by relative level of self-confidence (Robustness check).

		First switching point in EUR				Average switching point in EUR			
		Male subjects		**Female subjects**		**Male subjects**		**Female subjects**	
	Self-confi-dence[c]	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Man	Mean (Std.err.)	10.56 (0.94)	6.68 (0.72)	7.38 (0.57)	7.65 (1.17)	10.56 (0.94)	6.69 (0.66)	7.17 (0.53)	7.71 (0.88)
	Median	9.6	6.0	7.6	6.8	9.6	6.4	7.6	6.8
	Obs.	10	13	11	8	10	14	13	11
Selected-woman	Mean (Std.err.)	7.38 (0.55)	6.08 (0.68)	5.45 (0.60)	5.18 (0.76)	7.7 (0.62)	6.38 (0.65)	5.74 (0.59)	5.18 (0.65)
	Median	8.0	5.6	4.8	4.4	8.0	6.4	4.8	4.2
	Obs.	26	20	21	18	27	22	23	22
Difference in means [a]		3.18***	0.6	1.93***	2.47*	2.86**	0.30	1.43	2.53**
	[p-value]	[.005]	[.564]	[.046]	[.085]	[.020]	[.757]	[.113]	[.029]
Zero ATE[b]		$\chi^2(2)=8.889$[.002]***		$\chi^2(2)=8.641$ [.013]**		$\chi^2(2)=6.590$ [.037]**		$\chi^2(2)=8.682$ [.013]**	
Constant ATE[b]		$\chi^2(1)=3.076$[.079]*		$\chi^2(1)=0.112$ [.738]		$\chi^2(1)=3.062$ [.080]*		$\chi^2(1)=0.667$ [.414]	
Neutral	Mean (Std.err.)	7.65 (0.55)	8.48 (0.57)	8.82 (0.91)	8.24 (0.73)	7.65 (0.55)	8.38 (0.57)	8.68 (0.79)	8.31 (0.70)
	Median	7.6	8.4	9.6	8.4	7.6	8.4	9.0	8.4
	Obs.	15	7.02 (1.31)	17	15	15	17	20	18
Selected-neutral	Mean (Std.err.)	7.03 (0.73)		6.08 (0.65)	4.80 (0.56)	7.03 (0.73)	7.14 (1.22)	6.62 (0.75)	5.50 (0.64)
	Median	6.0	7.6	6.4	4.6	6.0	7.6	6.8	4.8
	Obs.	21	13	10	18	21	14	13	21
Difference in means[a]		0.62	1.46	2.74**	3.44***	0.62	1.24	2.06*	2.81***

Table B8 (*continued*)

	Self-confi-dence[c]	First switching point in EUR				Average switching point in EUR			
		Male subjects		**Female subjects**		**Male subjects**		**Female subjects**	
		High	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
	[*p*-value]	[.530]	[.291]	[.045]	[.001]	[.530]	[.339]	[.082]	[.005]
Zero ATE[b]		$\chi^2(2)=1.547$ [.464]		$\chi^2(2)=19.917$ [.000]***		$\chi^2(2)=1.315$ [.518]		$\chi^2(2)=12.311$ [.002]***	
Constant ATE[b]		$\chi^2(1)=0.248$ [.618]		$\chi^2(1)=0.231$[.631]		$\chi^2(1)=0.141$ [.707]		$\chi^2(1)=0.265$ [.607]	

***, **, *: Difference is significant on the 1, 5, 10 percent level. *p*-values in brackets.

[a] Two-sided *t* test.

[b] Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical) average effect for highly and less self-confident subjects.

[c] Self-confidence is high (low) if the belief about own score is above (below) the mean belief by gender.

Table B9

Performance evaluations by beliefs about own relative performance (Robustness check).

	Self-confi-dence[c]	First switching point in EUR				Average switching point in EUR			
		Male subjects		**Female subjects**		**Male subjects**		**Female subjects**	
		High	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Man	Mean	10.28	6.89	7.69	7.32	10.28	6.89	7.5	7.33
	(Std.err.)	(1.00)	(0.76)	(0.61)	(0.97)	(1.00)	(0.71)	(0.54)	(0.83)
	Median	9.2	6.0	8.0	6.8	9.2	6.4	7.8	6.8
	Obs.	10	13	9	10	10	14	12	12
Selected-Woman	Mean	6.51	7.18	4.89	5.70	6.88	7.37	5.03	5.78
	(Std.err.)	(0.53)	(0.73)	(0.69)	(0.64)	(0.63)	(0.68)	(0.67)	(0.57)
	Median	6	8.0	3.8	4.8	6.4	8.0	4.0	4.8
	Obs.	25	21	18	21	26	23	19	26
Difference in means[a]		3.77***	−0.29	2.80**	1.62	3.40***	−0.48	2.47**	1.55
	[p-value]	[.001]	[.794]	[.015]	[.167]	[.007]	[.638]	[.015]	[.134]
Zero ATE[b]		$\chi^2(2)=11.287$ [.004]***		$\chi^2(2)=11.212$ [.004]***		$\chi^2(2)=8.677$ [.013]**		$\chi^2(2)=10.630$ [.005]***	
Constant ATE[b]		$\chi^2(1)=6.922$ [.009]***		$\chi^2(1)=0.625$ [.429]		$\chi^2(1)=6.491$ [.011]**		$\chi^2(1)=0.482$ [.488]	
Neutral	Mean	7.40	8.65	9.67	7.11	7.20	8.78	9.41	7.39
	(Std.err.)	(0.61)	(0.48)	(0.66)	(0.93)	(0.60)	(0.47)	(0.60)	(0.86)
	Median	7.6	8.2	9.6	6.4	7.6	8.4	9.6	6.8
	Obs.	14	16	18	14	15	17	21	17
Selected-Neutral	Mean	6.02	8.02	5.06	5.46	6.02	8.07	5.65	6.21
	(Std.err.)	(0.51)	(1.19)	(0.62)	(0.64)	(0.51)	(1.12)	(0.70)	(0.70)
	Median	5.6	7.6	5.4	4.8	5.6	7.6	5.6	5.6
	Obs.	17	17	14	14	17	18	17	17
Difference in means[a]		1.38*	0.63	4.61***	1.66	1.18	0.71	3.76***	1.18
	[p-value]	[.093]	[.637]	[.000]	[.153]	[.146]	[.572]	[.000]	[.300]
Zero ATE [b]		$\chi^2(2)=3.213$ [.201]		$\chi^2(2)=27.931$ [.000]***		$\chi^2(2)=2.48$ [.280]		$\chi^2(2)=17.778$ [.000]***	
Constant ATE [b]		$\chi^2(1)=0.246$ [.620]		$\chi^2(1)=4.189$[.041]**		$\chi^2(1)=0.103$ [.748]		$\chi^2(1)=3.230$ [.072]*	

***, **, *: Difference is significant on the 1, 5, 10 percent level. *p*-values in brackets.

[a] Two-sided *t* test.

[b] Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical) average effect for highly and less self-confident subjects.

[c] Self-confidence classified as high (low) if beliefs about own minus the participant's score is above (below) the mean.

Table B10
Performance evaluations by absolute self-confidence (Robustness check).

		First switching point in EUR				Average switching point in EUR			
		Male subjects		**Female subjects**		**Male subjects**		**Female subjects**	
	Self-confi-dence[c]	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Man	Mean (Std.err.)	9.30	6.23	7.75	7.31	9.30	6.30	7.38	7.44
		(0.87)	(0.62)	(0.70)	(0.87)	(0.87)	(0.54)	(0.72)	(0.66)
	Median	9.2	5.6	8.4	6.8	9.2	6.2	8.0	6.8
	Obs.	16	7	8	11	16	8	9	15
Selected-Woman	Mean (Std.err.)	7.05	6.29	5.45	5.18	7.32	6.68	5.74	5.18
		(0.48)	(0.93)	(0.60)	(0.76)	(0.54)	(0.85)	(0.59)	(0.65)
	Median	7.6	5.8	4.8	4.4	7.6	7.8	4.8	4.2
	Obs.	32	14	21	18	33	16	23	22
Difference in means [a]		2.25**	−0.06	2.30**	2.13*	1.98*	−0.38	1.64	2.26**
	[p-Value]	[.018]	[.968]	[.040]	[.083]	[.051]	[.772]	[.127]	[.024]
Zero ATE [b]		$\chi^2(2)=4.958$ [.084]*		$\chi^2(2)=9.714$ [.008]***		$\chi^2(2)=3.774$ [.152]		$\chi^2(2)=9.065$ [.011]**	
Constant ATE [b]		$\chi^2(1)=2.403$ [.121]		$\chi^2(1)=0.013$[.907]		$\chi^2(1)=1.682$ [.102]		$\chi^2(1)=0.224$ [.636]	
Neutral	Mean (Std.err.)	8.13	7.80	8.82	8.24	7.98	8.23	8.68	8.31
		(0.49)	(0.31)	(0.91)	(0.73)	(0.49)	(0.50)	(0.79)	(0.70)
	Median	7.8	7.8	9.6	8.4	7.6	8.0	9.0	8.4
	Obs.	24	6	17	15	25	7	20	18
Selected-Neutral	Mean (Std.err.)	7.06	6.90	6.08	4.80	7.06	7.11	6.62	5.50
		(0.59)	(2.16)	(0.65)	(0.56)	(0.59)	(1.92)	(0.75)	(0.64)
	Median	6.8	5.6	6.4	4.6	6.8	7.6	6.8	4.8
	Obs.	26	8	10	18	26	9	13	21
Difference in means[a]		1.07	0.90	2.74**	3.44***	0.63	1.70	2.84*	2.97***
	[p-Value]	[.174]	[.729]	[.045]	[.001]	[.240]	[.625]	[.082]	[.005]
Zero ATE[b]		$\chi^2(2)=2.077$ [.354]		$\chi^2(2)=19.917$ [.000]***		$\chi^2(2)=1.736$ [.420]		$\chi^2(2)=12.311$ [.002]***	
Constant ATE[b]		$\chi^2(1)=0.06$ [.939]		$\chi^2(1)=0.231$[.631]		$\chi^2(1)=0.009$ [.925]		$\chi^2(1)=0.265$ [.607]	

***, **, *: Difference is significant on the 1, 5, 10 percent level. p-values in brackets.
[a] Two-sided t test.
[b] Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical) average effect for highly and less self-confident subjects.
[c] Self-confidence is high (low) if the subjects believes he/she is a top performer himself/herself.

Table B11
Performance evaluations by relative level of self-confidence (one-time switchers).

	Switching point		
Subjects:	**Male**	**Female**	**All**
Non-select TM * gender TM * self-confident * male			3.96*
			(2.14)
Non-select TM * gender TM * self-confident	4.70**	−.80	.15
	(2.00)	(2.40)	(1.94)
Non-select TM * self-confident * male			−.38
			(1.98)
Gender TM * self-confident * male			2.38
			(1.81)
Non-select TM * male			−2.30*
			(1.21)
Gender TM * male			−1.64
			(1.23)
Self-confident * male			−1.52
			(1.40)
Non-select TM * gender TM	−2.56*	−.64	−1.66
	(1.40)	(1.69)	(1.14)
Non-select TM * self-confident	−1.27	.77	−.49
	(1.29)	(1.63)	(1.61)
Gender TM * self-confident	.84	−.30	−1.17
	(1.35)	(1.44)	(1.44)
Male			2.28**
			(1.05)

Table B11 (*continued*)

Subjects:	Switching point		
	Male	**Female**	**All**
Non-select TM	2.14**	3.02***	3.89***
	(.99)	(1.06)	(1.00)
Gender TM	−.33	.03	.78
	(1.04)	(1.03)	(.99)
Self-confident	.41	.93	1.60
	(.98)	(1.05)	(1.03)
Age	.17	−.17*	.07
	(.15)	(.10)	(.13)
Constant	2.28	8.95***	2.91
	(3.69)	(2.53)	(3.29)
Obs.	108	96	204
R squared	.25	.29	.24

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Variable definition as in Tables B4 and B5. ***, **, *: Difference is significant on the 1, 5, 10 percent level. Standard errors in parentheses.

Table B12
Performance evaluations by relative level of self-confidence.

Subjects:	Average switching point		
	Male	**Female**	**All**
Non-select TM * gender TM * self-confident * male			4.04**
			(2.03)
Non-select TM * gender TM * self-confident	4.98**	−1.58	.14
	(1.90)	(2.15)	(1.79)
Non-select TM * self-confident * male			.09
			(1.76)
Gender TM * self-confident * male			2.04
			(1.74)
Non-select TM * male			−2.12*
			(1.12)
Gender TM * male			−1.17
			(1.12)
Self-confident * male			−1.77
			(1.38)
Non-select TM * gender TM	−2.83**	−.10	−1.43
	(1.32)	(1.46)	(1.06)
Non-select TM * self-confident	−1.25	.41	−.75
	(1.32)	(1.53)	(1.45)
Gender TM * self-confident	.67	.008	−.83
	(1.36)	(1.45)	(1.39)
Male			2.02**
			(1.03)
Non-select TM	2.11**	2.68***	3.49***
	(.97)	(1.04)	(.94)
Gender TM	−.16	−.37	.30
	(1.02)	(.99)	(.94)
Self-confident	.26	1.02	1.60*
	(1.00)	(1.09)	(1.05)
Age	.17	−.15*	.06
	(.15)	(.08)	(.12)
Constant	2.52	9.04***	3.66
	(3.63)	(2.09)	(3.00)
Obs.	113	116	229
R squared	.24	.24	.20

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Variable definition as in Tables B4 and B5. ***, **, *: Difference is significant on the 1, 5, 10 percent level. Standard errors in parentheses.

Table C1
Comparison of steady state results with two extreme cases ($g=4$).

0.500	0.380	0.277	0.194	0.132	0.090	Steady state
0.500	0.340	0.208	0.119	0.066	0.034	All female
0.500	0.403	0.311	0.231	0.166	0.115	All male

Appendix C. Practical implications—the glass ceiling

We close our numerical investigation by comparing our steady-state results with two extreme cases: a case where all promotion decisions are made entirely by men or all by women. The results for $g=4$ are shown in Table C1.

While all three cases are qualitatively similar[23], we see that the decrease in the number of female employees when moving up the hierarchy is most pronounced when all promotion decisions are made by women. In contrast, the proportion of females falls considerably slower than in the steady state when all promotion decisions are made by men. This exercise points out that the comparatively strong initial decrease in the proportion of females seen in Table 2 is driven significantly by the promotion decisions of females at intermediate hierarchy levels. Our findings are in line with previous literature investigating the gender composition of evaluation committees (Blau and DeVaro, 2007; Bagues and Esteve-Volart, 2010)

References

Anderson, L., Fryer, R., Holt, C., 2006. Discrimination: experimental evidence from psychology and economics. In: Rogers, W. (Ed.), Handbook on the Economics of Discrimination, Elgar, Cheltenham, UK, pp. 97.

Arrow, K., 1973. The theory of discrimination. In: Ashenfelter, O., Rees, A. (Eds.), Discrimination in Labor Markets, Princeton University Press, Princeton, pp. 3–33.

Bagues, M., Esteve-Volart, B., 2010. Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. Review of Economic Studies 77 (4), 1301–1328.

Bar-Hillel, M., 1980. The base-rate fallacy in probability judgments. Acta Psychologica 44 (3), 211–233.

Bauman, K., Geher, G., 2002. We think you agree: the detrimental impact of the false consensus effect on behavior. Current Psychology 21 (4), 293–318.

Becker, G., 1957. The Economics of Discrimination. University of Chicago Press, Chicago.

Blau, F., DeVaro, J., 2007. New evidence on gender differences in promotion rates: an empirical analysis of a sample of new hires. Industrial Relations: A Journal of Economy and Society 46 (3), 511–550.

Charness, G., Levin, D., 2005. When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. American Economic Review 95 (4), 1300–1309.

Crump, R., Hotz, V., Imbens, G., Mitnik, O., 2008. Nonparametric tests for treatment effect heterogeneity. Review of Economics and Statistics 15 (3), 389–405.

Dohmen, T., Falk, A., Huffman, D., Marklein, F., Sunde, U., 2009. Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample. Journal of Economic Behavior and Organization 72 (3), 903–915.

El-Gamal, M., Grether, D., 1995. Are people Bayesian? Uncovering behavioral strategies. Journal of the American Statistical Association 90 (432), 1137–1145.

Falk, A., D. Huffman, U. Sunde, 2006. "Self-confidence and search." Institute for the Study of Labor (IZA); University of St. Gallen Working Paper.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2), 171–178.

Greiner, B., 2003. An online recruitment system for economic experiments. In: Forschung und wissenschaftliches Rechnen, Kremer, K., Machoeds., V., (Eds.) GWDG Bericht 63. Göttingen: Ges. für Wiss. Datenverarbeitung. pp. 79–93.

Grether, D., 1992. Testing Bayes rule and the representativeness heuristic: some experimental evidence. Journal of Economic Behavior and Organization 17 (1), 31–57.

Kahneman, D., Tversky, A., 1972. Subjective probability: a judgment of representativeness. Cognitive Psychology 3 (3), 430–454.

Lyon, D., Slovic, P., 1976. Dominance of accuracy information and neglect of base rates in probability estimation. Acta Psychologica 40 (4), 287–298.

Möbius, M., Niederle, M., Niehaus, P., Rosenblat, T., 2011. Managing self-confidence: theory and experimental evidence. National Bureau of Economic Research.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much?. Quarterly Journal of Economics 122 (3), 1067–1101.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., Richardson, C., 1995. A redrawn Vandenberg and Kuse mental rotations test-different versions and factors that affect performance. Brain and Cognition 28 (1), 39–58.

Phelps, E., 1972. The statistical theory of racism and sexism. American Economic Review 62 (4), 659–661.

Reuben, E., P. Sapienza, and L. Zingales. 2010. The Glass Ceiling in Experimental Markets Working Paper.

Ross, L., Greene, D., House, P., 1977. The 'false consensus effect': an egocentric bias in social perception and attribution processes. Journal of Experimental Social Psychology 13 (3), 279–301.

Tversky, A., Kahneman, D., 1971. Belief in the law of small numbers. Psychological Bulletin 76 (2), 105–110.

Vandenberg, S., Kuse, A., 1978. Mental rotation, a group test of three-dimensional spatial visualization. Perceptual and Motor Skills 47 (2), 599–604.

[23] This qualitative similarity is also a robustness check for our fixed-point procedure.