

# Negative Side Effects of Affirmative Action

## How Quotas Lead to Distortions in Performance Evaluation

Lea M. Petters\* Marina Schröder†

This version: June 18, 2019

### *Abstract*

We study the impact of affirmative action on peer-review behavior. We conduct a real-effort tournament experiment, in which prizes are awarded according to peer-review results. Between treatments, we vary whether or not a quota is implemented and whether or not affirmed individuals face ex-ante procedural disadvantages. We find that quotas lead to distorted peer-reviews, where affirmed individuals receive significantly less favorable reviews than non-affirmed peers with similar performance. Distortions in peer-reviews are robust to changes in ex-ante procedural fairness and seem to be driven by enhanced competition among affirmed individuals.

**Key Words:** Affirmative action, Quota, Peer-review, Creative work, Real effort, Procedural fairness

**JEL Classification:** C92, J33, J71, M51

---

\* University of Cologne, Seminar of Personnel Economics and Human Resource Management, Albertus-Magnus-Platz, 50923 Cologne, Germany, phone: +49 221 470 6312, Email: lea.petters@uni-koeln.de.

† University of Cologne, Seminar of Personnel Economics and Human Resource Management, Albertus-Magnus-Platz, 50923 Cologne, Germany, phone: +49 221 470 6939, Email: marina.schroeder@uni-koeln.de.

We thank the Center for Social and Economic Behavior (C-SEB) through the C-SEB Gender Research Grant and the German Research Foundation (DFG) through priority program SPP 1764 (SL 46/2-1) and the research unit “Design and Behavior – Economic Engineering of Firms and Markets” (FOR 1371) for financial support. We thank the audience at the ESA World Meeting 2016 in Jerusalem, ESA European Meeting in Bergen, IMEBESS 2017 in Barcelona, the Research Seminar in Applied Microeconomics in Cologne, the C-SEB Workshop at the University of Cologne, and the Gender Economics Workshop at the DIW Berlin, and Stefano Balietti, Uri Gneezy, Bernd Irlenbusch, Andreas Leibbrandt, Ernesto Reuben, Dirk Sliwka, and Susanne Steffes for helpful comments and suggestions. We thank Tobias Danzeisen and Ruth Neeßen for excellent support in programming and conducting the experiment.

# 1. Introduction

Despite numerous endeavours to promote the career advancement of females and ethnic minorities, these groups are still underrepresented in leading positions in management (European Commission 2016; 2020 Women on Boards 2018; Beech et al. 2017) and academia (Lundberg and Stearns 2019; American Economic Association 2018). Research reveals two important channels that explain the underrepresentation of females and ethnic minorities in leading positions: differences in career-relevant behavior (Babcock et al. 2017; Coffman 2014; Croson and Gneezy 2009; Niederle and Versterlund 2007) and discrimination (Beaurain and Masclet 2016; Bertrand and Duflo 2016; Mengel, Sauermann, and Zölitz 2019; Sarsons 2017). As a mean to counteract this underrepresentation, some countries and organizations implement affirmative action policies in the form of quota regulations.<sup>1</sup> Supporting the introduction of such intervention, several studies show that quotas are effective at reducing differences in career-relevant behavior and that quotas therefore increase the representation of the affirmed group without harming efficiency (Balafoutas and Sutter 2012; Niederle, Segal, and Vesterlund 2013; Banerjee et al. 2013).

While designed to promote an important goal, affirmative action may also entail negative effects. In this paper, we provide evidence for negative side effects of quota interventions on subjective peer-reviews. Distortions in subjective performance evaluation are considered a crucial facilitator for discrimination (Borgida and Fiske 2008; Nieva and Gutek 1980). As peer-reviews are especially relevant for career success in management and academia (Bracken 2001; Edwards and Ewen 1996; Johnson 2004), distorted evaluations may have long-term detrimental effects.<sup>2</sup> Our results show that quotas lead to distortions in peer-reviews of the affirmed group and therefore may actually increase discrimination against the group that is supposed to benefit from the quota. Furthermore, the anticipation of distortions in performance evaluation may mitigate the positive effect of quotas on career-relevant behavior of the affirmed group (Leibbrandt, Wang, and Foo 2018; Leibbrandt and List 2018). Finally, distortions in peer-review may affect the self-image and actual performance of affirmed individuals in an unfavorable way (Heilman and Alcott 2001; Turner and Pratkanis 1993).

Previous research on the effect of affirmative action on performance evaluation is mixed and has focused on evaluations provided by bystanders or superiors. On the one hand, increased exposure to competent affirmed individuals due to the introduction of quotas is shown to reduce biases in performance

---

<sup>1</sup> Countries that have passed laws to implement gender quotas for management positions include Norway, Spain, Italy, Belgium, France, and Germany. Sowell (2004) and Bagde, Epple, and Taylor (2015) present examples for quotas based on ethnic background. Wallon, Bendiscioli, and Garfinkel (2015) provide an overview for the use of quota interventions in academia.

<sup>2</sup> In management, peer-reviewing is widespread. The majority of companies listed in the Fortune 500 use peer-reviews as a tool for subjective performance evaluation of management positions (Bracken 2001; Edwards and Ewen 1996). In academia, peer-reviews are relevant for placement and tenure decisions, publication success and research funding.

evaluation (Beaman et al. 2009; Baskaran and Hessami 2018). On the other hand, affirmative action is shown to have a negative effect on the evaluation of successful affirmed individuals because success is overly ascribed to affirmative action rather than ability (Bijkerk et al. 2018; Heilman, Block, and Lucas 1992; Heilman, Block, and Stathatos 1997; Gürtler and Gürtler 2015). In a recent closely related paper on the effect of quotas on peer-reviews in a gender context, Leibbrandt, Wang, and Foo (2018) show that women are significantly more likely to be a victim of sabotage (intentional misreporting of objective performance measures through peers) whenever a female quota is implemented. While Leibbrandt, Wang, and Foo (2018) provide first indications that quotas may have an impact on peer-reviewing behavior, we know little about the mechanism driving this effect or about the generalizability of this result beyond the gender context.

Quotas can impact peer-reviewing behavior for different reasons. First, affirmative action interventions substantially change the competitive structure of a tournament (Chowdhury and Gürtler 2015; Franke 2012; Calsamiglia, Franke, and Rey-Biel 2013; Schotter and Weigelt 1992; Holzer and Neumark 2000). Enhanced competition among affirmed individuals due to a quota is likely to lead to distortions in peer-reviews as affirmed individuals have an incentive to provide less favorable peer-reviews to affirmed peers.<sup>3</sup> Second, quotas increase the winning probabilities of affirmed individuals at the cost of non-affirmed peers. Inequity averse individuals (affirmed and non-affirmed) may react to a quota by providing distorted peer-reviews favoring non-affirmed individuals to counteract procedural unfairness (Cappelen et al. 2007; Konow 1996).

In our experiment, we randomly assign individuals to affirmed or non-affirmed types.<sup>4</sup> Within groups of four, participants compete for two prizes each by performing a creative real-effort task (Laske and Schröder 2016).<sup>5</sup> Prizes in the tournament are awarded according to subjective peer-reviews and - depending on the treatment - a quota. In treatments without a quota, the two prizes are rewarded to the two (out of four) participants with the highest score in the peer-review. In treatments with a quota, at least one of the two prizes has to be awarded to one of the two randomly determined affirmed type participants. As previous research reveals that the effect of affirmative action depends on its perceived justification (Balafoutas, Davis, and Sutter 2016; Harrison et al. 2006; Ip, Leibbrandt, and Vecchi

---

<sup>3</sup> Subjective peer-reviews are prone to - conscious or unconscious - biases. In settings without a quota, peer-reviews are shown to be sensitive to changes in the incentive structure (see e.g. Balmelli, Goldstone, and Helbing 2016; Carpenter, Matthews, and Schirm 2010; Rosaz and Villeval 2012). Harbring et al. (2007) show that sabotage behavior in a contest is affected by the symmetry of the tournament.

<sup>4</sup> Random assignment of a quota is an important difference between our experimental study and most previous experimental studies on the effect of quotas (see e.g. Leibbrandt, Wang, and Foo 2018; Niederle, Segal, and Vesterlund 2013). Randomly assigning the quota provides an advantage by allowing us to cleanly induce unequal opportunities and thus study the effect of justification of a quota. Furthermore, due to the random assignment of a quota, we can rule out that our findings are specific to certain groups, i.e., females or non-whites. Thus, our findings can also be applied to other types of quotas, i.e., quotas based on race or caste.

<sup>5</sup> The use of a creative task has two important advantages. First, the use of subjective performance evaluations is very natural in this context because by definition creative performance cannot be quantified through objective measures. Second, creative work is especially relevant in leading positions in management and in academia. Thus, creative work is highly relevant for the type of work in which quotas are often implemented.

(forthcoming)), we conduct our experiment in two different settings. In the ex-ante equal setting, affirmed and non-affirmed individuals face the same procedure for working on the task. In the ex-ante unequal setting, affirmed individuals face procedural disadvantages that provide a possible justification for the introduction of a quota.

We find evidence for substantial biases in peer-reviews due to the introduction of a quota. First, quotas have an impact on the average level of peer-reviews provided. The observed level effects vary depending on the setting (ex-ante equal or ex-ante unequal) and thus seem to be related to the perceived procedural fairness of the quota. Second, quotas lead to substantial distortions in peer-reviews, such that under a quota, affirmed individuals receive significantly less favorable peer-reviews compared to non-affirmed peers with similar performance according to an independent measure. Importantly, we show that these distortions in peer-reviews are unrelated to procedural inequalities and are driven by reviews provided by affirmed individuals. Thus, it seems that distortions in peer-reviews are the result of enhanced competition among affirmed individuals due to the introduction of a quota. In a subsequent dictator game (Forsythe et al. 1994), we find that facing a quota in the previous tournament experiment significantly reduces subsequent altruistic behavior among affirmed individuals. Thus, we provide evidence for behavioral spillovers of quotas beyond the context in which they are implemented.

## **2. Experimental Design**

In our experiment, participants face a tournament setting in which groups of four participants compete for two prizes. At the beginning of the experiment, we randomly assign participants as one of two types (yellow or green). Each group consists of two yellow and two green type participants. Each round of the tournament consists of two stages: a working stage and a peer-review stage. In the working stage, participants perform a creative real-effort task. The task consists of illustrating predefined objects using a given set of materials. It is a modified version of the task introduced by Laske and Schröder (2016). Participants receive a set of materials: one string, two O-rings, four wooden sticks, and twelve colored glass pebbles (see Figure 1) and are asked to use these materials to illustrate specific objects, i.e. eyeglasses, a flower, and a car. Participants are instructed to take pictures of their illustrations using a special software and a pre-installed webcam. See Figure 2 for examples of pictures of the illustrations and Appendix 1 for the experimental instructions. The time available in the working stage is restricted and depends on the treatment and the type of the participant (see Table 1 for an overview of our treatments). Within the limited time frame, participants can take as many pictures as they want. Once the time is up, participants choose one of these pictures to be payoff-relevant. All other pictures are deleted and not payoff-relevant.

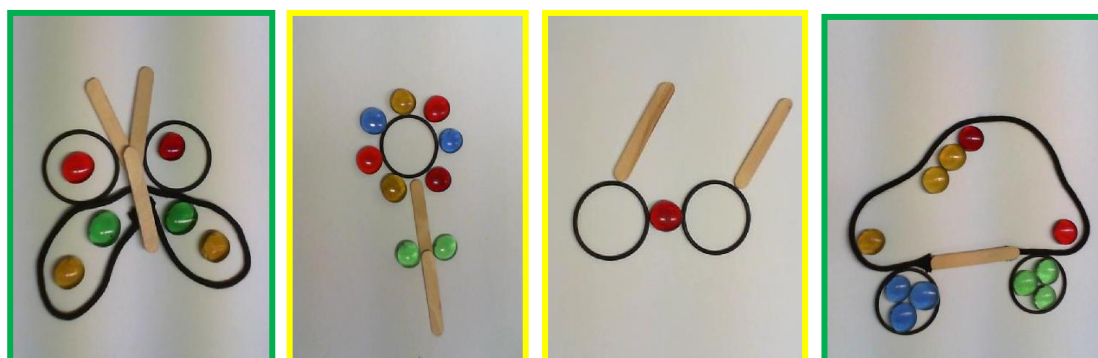


**Figure 1: Set of materials**

In the peer-review stage of each tournament round, participants see the pictures of the illustrations created by their group members (including their own illustration) and are asked to rate the illustrations of their peers (participants did not rate their own illustrations) on a scale from 0.0 to 10.0 (0.0 being the worst rating and 10.0 being the best rating). Thus, in the peer-review stage, the illustrations of all participants are evaluated by the three other participants of each group. Prizes are awarded according to the mean rating from this peer-review and, if applicable, a quota. We conduct three rounds of the tournament without feedback and with random rematching within matching groups of eight participants between rounds.

Between treatments, we vary whether or not a quota is implemented. In treatments without a quota, the two participants with the highest and the second highest peer-review receive a prize independent of their type. In treatments with a quota, at least one of the two prizes is awarded to a participant of the affirmed type. Thus, in treatments involving a quota, the participant among the affirmed types with the highest mean rating from the peer-review receives a prize for sure (even if this participant is not among the two participants with the highest mean ratings) and the participant among the remaining three participants of either type with the highest mean rating from the peer-review receives a prize.

We consider two different settings in which quotas are implemented. In the ex-ante equal setting, all participants face the same procedure absent of a quota, which means that all types have five minutes to work on the task in each round. In the ex-ante unequal setting, one type faces procedural disadvantages by having a reduced working time of only 2 minutes and 30 seconds.



**Figure 2: Examples of illustrations**

When a quota is implemented in this setting, the type with the shorter working time is also affirmed. Table 1 summarizes the treatments of our experiment. In the *Baseline\_equal* treatment all participants face the same procedure and no quota is implemented. In the *Quota\_equal* treatment, we implement a quota in a setting where all participants face the same procedure. In the *Baseline\_unequal* treatment, one type faces reduced working time, but no quota is implemented. In the *Quota\_unequal* treatment, the type that faces reduced working time is also affirmed by a quota.

The quota is implemented according to the randomly assigned type (yellow or green) and thus independent of any characteristics of the participants. In all treatments, we randomized whether green or yellow individuals are affirmed and/or face procedural disadvantages. In the peer-review stage, the pictures are shown in a randomized order. They have a colored frame, which indicates the type of the ratee (see frames in example pictures provided in Figure 2). Thus, the types - but not the identity of the other participants - are common information.

Participants are paid for only one randomly determined round of the main experiment. The two winners of the tournament in this round receive a prize of 16 euros each while the other participants receive zero for their performance in the task. In addition to payment for performance in the task, all participants receive a show-up fee of 4 euros and can earn up to 2 euros in the subsequent dictator experiment. At the end of each experimental session, we distributed a pen and paper questionnaire (see Appendix 2 for the full questionnaire), in which we ask about demographics as well as perceived fairness of the experimental procedure, which participants should rate on a scale from 1 (very unfair) to 5 (very fair). We conducted the experiments in April and October 2016 and May 2017 at the Cologne Laboratory for Economic Research (CLER). Overall, 632 subjects participated in our experiment and we ran 40 experimental sessions. We used Java to program our experiments and recruited the participants through the online recruitment software ORSEE (Greiner 2015).

**Table 1: Treatments**

	Ex-ante equal	Ex-ante unequal
No quota	<p><b><i>Baseline_equal</i></b>  No quota  Equal working time for all  n=160; N=20</p>	<p><b><i>Baseline_unequal</i></b>  No quota  Disadvantaged type has reduced working time  n=160; N=20</p>
Quota	<p><b><i>Quota_equal</i></b>  At least one prize awarded to affirmed type  Equal working time for all  n=160; N=20</p>	<p><b><i>Quota_unequal</i></b>  At least one prize awarded to affirmed &amp; disadvantaged type  Affirmed &amp; disadvantaged type has reduced working time  n=152; N=19<sup>6</sup></p>

n indicates the total number of participants in each treatment.

N indicates the number of independent observations, i.e., matching groups of 8 participants, in each treatment.

Our treatment interventions can have an effect on peer-reviewing behavior, but it can also affect performance per se. To be able to distinguish between these two effects, we elicit a performance measure that is independent of our treatment interventions. We conduct an online experiment, in which we ask a total of 400 independent raters to evaluate the illustrations from the experiment on the same scale as in the laboratory experiment (0.0 to 10.0). To avoid overload of the raters, each rater evaluates a subset of 64 pictures. In the online experiment, each screen shown to the raters displays four pictures from one group in one round. The composition of pictures on one screen shown to the independent raters was identical to that shown to the participants of the experiment. The pictures also have the yellow and green frames as in the laboratory experiment. One important difference to the main experiment is that the evaluators are blind to the treatment and do not know what the frames mean. Each set of pictures is evaluated by ten different raters. The average of these ten evaluations constitutes our independent quality measure. Participants in the online experiment receive a fixed payment of 2 euros and can earn an additional bonus of up to 4 euros, which is awarded according to the quadratic deviation from the mean evaluation for one randomly chosen picture (we follow the procedure suggested by Selten (1998)). We programmed this experiment using the online survey tool SoSci Survey (Leiner 2014). Online raters were also recruited through the online recruitment software ORSEE (Greiner 2015) using the same subject pool as in the main experiment but ensuring that evaluators did not participate in any previous related experiments.

<sup>6</sup> One group had to be eliminated from the analysis because one subject from this group did not pass the control questions.

### 3. Results

#### 3.1 Biases in Peer-reviews

Across all treatments, the average peer-review is 2.59 points, while the average score on the independent evaluation is equal to 6.00 (see Appendix 3 for summary statistics). For all treatments and types, we find that peer-reviews are significantly less favorable as compared to the independent ratings (pairwise Wilcoxon rank-sum test,  $p < 0.01$  for all types and treatments), but significantly larger than zero (Wilcoxon rank-sum test,  $p < 0.01$  for all types and treatments).<sup>7</sup> Thus, we find evidence for a substantial difference between the independent measure and the peer-review.<sup>8</sup>

To analyze biases in peer-reviews that are due to the introduction of a quota, we provide regression analysis results in Table 3. In specifications (1) and (2), we focus on the ex-ante equal setting and in specifications (3) and (4), we focus on the ex-ante unequal setting.<sup>9</sup> In specification (1) we only include a quota dummy and a dummy for affirmed ratees (receiver of an evaluation). The coefficient of the quota dummy is informative of the effect of a quota on the overall level of peer-reviews. The coefficient of the affirmed ratee dummy is informative of distortions in peer-reviews, i.e., systematic differences of peer-reviews provided depending on the type of the ratee.

In the ex-ante equal setting, we find a slight but insignificant negative level effect. Additionally, we find a significant and negative coefficient of the affirmed ratee dummy. Affirmed ratees receive around 0.6 points less favorable evaluations, which amounts to 21% of the average evaluation in the baseline. This provides evidence for a significant distortion effect, where affirmed types receive less favorable peer-reviews compared to non-affirmed peers. To better understand the extent to which this distortion is due to changes in the performance of affirmed types, we include the independent performance measure as an additional control in specification (2).<sup>10</sup> Introducing this control reduces the distortion effect slightly, but the coefficient for affirmed ratees is still large and statistically significant.

In the ex-ante unequal setting (specifications (3) and (4)), we find that the introduction of a quota has a positive and significant effect on the overall level of peer-reviews provided. That is, we find that overall peer-reviews provided are about 0.6 points more favorable and thus, closer to the independent rating, when a quota is implemented. As in the ex-ante equal setting, we find evidence for a significant

---

<sup>7</sup> For the non-parametric analysis in this paper, we use mean values for the matching groups of eight participants and over all rounds of the experiment. We always report p-values for two-sided tests.

<sup>8</sup> In order to address the concern that our independent online raters have no experience with the task that they evaluate, we asked a subset of our participants in the laboratory experiment to rate illustrations created in different sessions after they have completed the main experiment. We only elicited this measure for the ex-ante equal setting. In this setting, however, we can show that the main results presented in this paper are robust to using this alternative performance measure as a control variable. Analysis using experienced lab raters can be found in Appendix 4.

<sup>9</sup> In Appendix 5 we provide an additional specification which analyzes both settings in one model.

<sup>10</sup> In Appendix 6 we provide an analysis of the effect of our treatment interventions on performance. We do not find evidence for a significant effect of quotas on the performance as measured through independent ratings. However, we do find that procedural disadvantages induced by reduced working time are in fact relevant and have a negative effect on performance.



distortion effect, where affirmed types receive less favorable peer-reviews compared to non-affirmed peers. Affirmed ratees receive evaluations that are around 0.6 points less favorable compared to non-affirmed ratees. This distortion amounts to around 25% less favorable evaluations compared to the corresponding reference group in the baseline. Interestingly, we do not find a significant effect of being disadvantaged on the peer-reviews received. In column (4), we additionally control for performance. If anything, including this control explains a very small part of the distortion in peer-reviews, since the coefficient of affirmed remains large and highly significant. Again, we find no significant difference in peer-reviews provided to disadvantaged types compared to non-disadvantaged peers.

**Table 2: Regression analysis peer-reviews provided**

<i>Dependent variable:</i> Peer-review	Ex-ante equal		Ex-ante unequal	
	(1)	(2)	(3)	(4)
Quota	-0.130 (0.217)	-0.036 (0.240)	0.634** (0.283)	0.645** (0.300)
Affirmed ratee	-0.562*** (0.153)	-0.502*** (0.143)	-0.632*** (0.229)	-0.544*** (0.199)
Disadvantaged ratee			-0.011 (0.135)	0.178 (0.118)
Independent rating		0.341*** (0.026)		0.382*** (0.026)
Constant	2.884*** (0.157)	0.734*** (0.214)	2.383*** (0.221)	0.012 (0.247)
Observations	2,880	2,880	2,808	2,808
Number of participants	320	320	312	312
Number of groups	40	40	39	39

Two-way error component linear model, allowing for creator and evaluator random effects. The dependent variable is peer-review received. Independent variables: Quota (dummy equal to one in treatments with a quota), Affirmed ratee (dummy equal to one for affirmed ratees in the treatments with a quota), Disadvantaged ratee (dummy equal to one for disadvantaged ratees in all treatments involving less working time for disadvantaged type), Independent Rating (continuous variable with the mean evaluation of independent raters who are blind to treatments)

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

To understand the source of these biases in peer-reviews, we analyze the behavior of affirmed and non-affirmed raters (sender of an evaluation) separately. Table 3 provides results using the same regression procedure as in specifications (2) and (4) of Table 2, but splits the sample depending on the type of the rater. In the ex-ante equal setting, we use all observations from the baseline as a reference group for both the affirmed and the non-affirmed types. Specification (1) in Table 3 presents the results for affirmed raters in the ex-ante equal setting. While we find no effect of quotas on the overall level of peer-reviews provided by affirmed raters, we find evidence for a strong distortion effect where affirmed raters provide significantly less favorable peer-reviews to affirmed peers, i.e., around 0.6 points less favorable

evaluations. Specification (2) presents the results for non-affirmed raters in the ex-ante equal setting. We find evidence for a significant level effect, i.e., non-affirmed raters provide around 0.8 points less favorable peer-reviews overall whenever quotas are implemented. However, we do not find evidence for a significant distortion effect of non-affirmed raters.

Specifications (3) and (4) provide the results for the ex-ante unequal setting. In specification (3), we focus on affirmed raters. With respect to the level effect, we find that affirmed raters provide significantly more favorable peer-reviews, i.e., around 1.0 point more favorable evaluations compared to the baseline, whenever a quota is implemented. We also find evidence for a large and significant distortion effect, where affirmed raters provide of around 0.7 points less favorable peer-reviews to other affirmed peers as compared to non-affirmed peers. As presented in specification (4), we do not find evidence for any systematic biases in peer-reviews provided by non-affirmed raters when a quota is introduced in the ex-ante unequal setting.

**Table 3: Regression analysis peer-reviews provided by rater type**

<i>Dependent variable:</i> Peer-review	Ex-ante equal		Ex-ante unequal	
	(1) Affirmed rater	(2) Non-affirmed rater	(3) Affirmed rater	(4) Non-affirmed rater
Quota	0.217 (0.306)	-0.806*** (0.282)	1.028*** (0.344)	-0.094 (0.407)
Affirmed ratee	-0.595*** (0.190)	0.137 (0.088)	-0.701*** (0.198)	0.111 (0.178)
Disadvantaged ratee			0.003 (0.118)	0.072 (0.145)
Independent rating	0.346*** (0.032)	0.339*** (0.027)	0.347*** (0.044)	0.400*** (0.036)
Constant	0.690*** (0.236)	0.747*** (0.210)	0.101 (0.333)	0.179 (0.300)
Observations	2,160	2,160	1,404	1,404
Number of participants	320	320	312	312
Number of groups	40	40	39	39

Two-way error component linear model, allowing for creator and evaluator random effects. Separate models for affirmed and non-affirmed raters. The dependent variable is peer-review received. Independent variables: Quota (dummy equal to one in treatments with a quota), Affirmed ratee (dummy equal to one for affirmed ratees in the treatments with a quota), Disadvantaged ratee (dummy equal to one for disadvantaged ratees in all treatments involving less working time for disadvantaged type), Independent Rating (continuous variable with the mean evaluation of independent raters who are blind to treatments). In the ex-ante equal setting, all raters from the *Baseline\_equal* serve as a reference group for both affirmed and non-affirmed raters. In the ex-ante unequal, disadvantaged raters from the *Baseline\_unequal* treatment serve as the reference group for affirmed raters, while the non-disadvantaged individuals from the *Baseline\_unequal* treatment serve as reference group for non-affirmed raters in the *Quota\_unequal* treatment.

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Thus, it seems that level effects depend on the setting in which a quota is implemented, while distortion effects are fully driven by reviews provided by affirmed raters and seem to be independent of the setting. As a robustness check for this interpretation, we provide the results from regression analyses in which we additionally control for perceived procedural fairness in Appendix 7. We find that adding this control substantially reduces observed level effects but has basically no effect on observed distortion effects.<sup>11</sup>

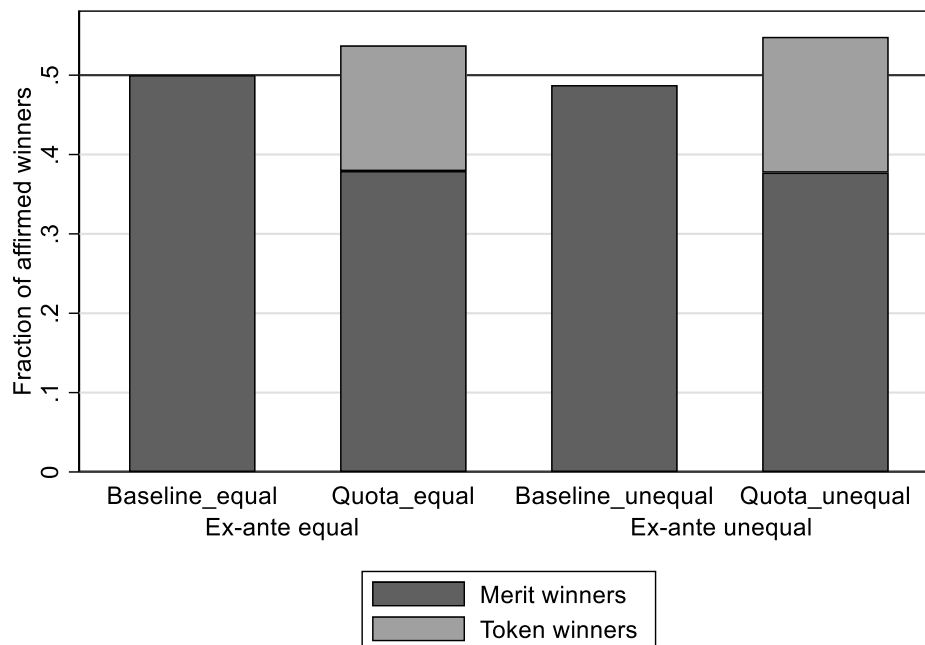
### 3.2 Tokenization of Affirmed Winners

In Figure 3, we display the fraction of tournament winners from the affirmed group by treatment. We additionally differentiate between “merit winners” – winners of the affirmed type who are among the two participants with the highest peer-review – and “token winners” – affirmed winners who are not among the two performers with the highest peer-review, but win because of the quota intervention.

We find that in both the ex-ante equal and the ex-ante unequal settings, the introduction of the quota significantly increases the overall number of tournament winners from the affirmed type (Mann-Whitney U-test,  $p < 0.01$  and  $p = 0.08$ , correspondingly). In both the *Baseline\_equal* and the *Baseline\_unequal* treatments, by design, all winners are “merit winners”. In the *Quota\_equal* and the *Quota\_unequal* treatments, in absence of side-effects of quotas, the fraction of “merit winners” would be unaffected by the introduction of a quota and the fraction of “token winners” would correspond to the increase in the representation of affirmed types among tournament winners. However, in our experiment we observe substantial distortions in peer-reviews. Thus, we find that quotas lead to a decrease in the fraction of affirmed “merit winners” (Whitney U-test,  $p \leq 0.01$  for both settings). As a consequence, the fraction of “token winners” is larger than the actual increase in the representation of affirmed types among tournament winners. Therefore affirmed types appear to be in the need of a quota to win the tournament, whereas a large fraction of these “token winners” would also have been among the winners without a quota intervention.

---

<sup>11</sup> In Appendix 8, we provide further analysis splitting the sample into male and female raters. We find that distortion effects are independent of gender. With respect to level effects, we observe some gender differences, where affirmed females raters provide more favorable peer-reviews when a quota is implemented.



In the ex-ante equal setting, one random type from the *Baseline\_equal* serves as reference group for affirmed types in the *Quota\_equal* treatment. In the ex-ante unequal setting, disadvantaged types from the *Baseline\_unequal* treatment serve as reference group for affirmed and disadvantaged types in the *Quota\_unequal* treatment.

**Figure 3: Fraction of affirmed winners**

### 3.3 Spillover Effects of Quotas

Lastly, we are interested in spillover effects of quotas on behavior in a dictator game that does not involve any quota regulation. Therefore, we conduct a dictator experiment (Forsythe et al. 1994) after the main experiment. Besides the fact that we inform dictators about the receiver's type from the previous experiment, this experiment is unrelated to the main experiment. The instructions for this experiment are given after completion of the main experiment (see Appendix 9 for the experimental instructions). In this experiment, dictators allocate 2 euros between themselves and an anonymous recipient. Dictators are informed about the type of the recipient (yellow or green) from the previous tournament experiment. We repeat the dictator game four times with random rematching, so that each participant acts twice as a dictator and twice as a recipient. In both roles (dictator and recipient), each participant is matched to both a yellow and a green type player. We randomized the sequence of the four rounds. Only one of the rounds is randomly chosen for payment, and no feedback is given between the rounds.

In Table 4, we show the results from a linear regression analysis with the amount sent in the dictator game as the dependent variable and a quota dummy and dummies for affirmed and if applicable, disadvantaged receiver as independent variables. Specification (1) and (2) show the results for the ex-

ante equal setting, and specification (3) and (4) for the ex-ante unequal setting. For each setting, we analyze dictator behavior for the previously affirmed and non-affirmed group separately.

**Table 4: Regression analysis dictator game experiment by dictator type**

<i>Dependent variable:</i> Amount sent in dictator game	Ex-ante equal		Ex-ante unequal	
	(1) Affirmed dictator	(2) Non-affirmed dictator	(3) Affirmed dictator	(4) Non-affirmed dictator
Quota	0.081 (0.060)	-0.029 (0.055)	0.078 (0.058)	0.010 (0.064)
Affirmed receiver	-0.067** (0.027)	-0.106*** (0.027)	-0.064* (0.033)	-0.017 (0.037)
Disadvantaged receiver			0.088*** (0.026)	-0.004 (0.022)
Constant	0.324*** (0.032)	0.324*** (0.032)	0.211*** (0.040)	0.295*** (0.045)
Observations	480	480	312	312
Number of participants	240	240	156	156
Number of groups	40	40	39	39

Ordinary-least squares linear model. Separate models for affirmed and non-affirmed dictators. The dependent variable is the amount sent in dictator game. Independent variables: Quota (dummy equal to one if a quota was implemented in the previous experiment), Affirmed receiver (dummy equal to one if receiver was affirmed in the previous experiment), Disadvantaged receiver (dummy equal to one if receiver was disadvantaged in the previous experiment). In the ex-ante equal setting, all individuals from the *Baseline\_equal* serve as a reference group for both affirmed and non-affirmed dictators. In the ex-ante unequal, the dictators that were disadvantaged in the previous *Baseline\_unequal* treatment serve as the reference group for the affirmed dictators, while the dictators that were non-disadvantaged in the previous experiment serve as reference group for non-affirmed types in the *Quota\_unequal* treatment.

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Throughout all specifications, we do not find any significant effect of facing a quota in the previous experiment on the overall level of giving. However, we do observe distortion effects where participants, who were affirmed in the previous experiment, receive significantly lower contributions in the subsequent dictator game. Independent of the setting, we find that a quota in the previous experiment leads to lower giving by dictators, who were previously affirmed, to previously affirmed receivers of around 0.07 and 0.06 euros respectively. In both treatments, this amounts to 22% less compared to the average in the corresponding baseline. For dictators who were non-affirmed in the previous experiment, we find that distortions in dictator giving depend on the setting of the previous experiment. Dictators who faced the ex-ante equal setting in the previous experiment, give around 0.11 euros less to receivers who were affirmed in the previous experiment as compared to receivers who were non-affirmed, i.e., around 34% less compared to the *Baseline\_equal* treatment. Dictators who

faced the ex-ante unequal setting in the previous experiment, do not discriminate based on the type of the receiver in the previous experiment.

## **4. Discussion and Conclusion**

In this paper, we provide evidence for biases in subjective performance evaluations which arise due to the introduction of a quota. We find that quotas have a significant effect on the overall level of peer-reviews provided and that this effect varies with the setting in which quotas are implemented. In our experimental study, the average level of peer-review does not have an impact on tournament outcomes. Thus, level effects do not counteract procedural inequalities arising due to the implementation of a quota. Nevertheless, we observe that the effect of a quota on the level of peer-reviews provided is sensitive to the setting in which quotas are implemented. We can show that this difference is mainly driven by differences in perceived procedural fairness due to the introduction of a quota. One explanation for this finding is that costs of providing biased reviews depend on perceived procedural fairness. This interpretation is in line with previous findings on the effect of perceived fairness on unethical behavior (see e.g. Greenberg 1990; Houser, Vetter, and Winter 2012; Schweitzer and Gibson 2008). In real-life work situations, such level effects may have additional motivational implications (Heilman and Alcott 2001; Unzueta, Gutiérrez, and Ghavami 2010) and may be problematic whenever peer-reviews elicited under a quota regime are compared to reviews elicited absent of a quota regime.

We additionally show that quotas lead to systematic distortions in peer-reviews, where affirmed individuals receive less favorable peer-reviews as compared to non-affirmed individuals whenever quotas are implemented. These distortions seem to be unrelated to ex-ante procedural fairness. Given that distortions are fully driven by peer-reviews provided by affirmed types, it seems that distortions are the consequence of enhanced competition among affirmed individuals whenever quotas are implemented. While a quota compensates for the potential distortions in prize assignment due to biased peer-reviews, quotas may have a long-term negative impact on the career chances of affirmed individuals. Distortions make affirmed types appear to be less able and in the need of a quota even when they are not. As peer-reviews are widely used in domains where objective performance measures are lacking, such distortions can have far-reaching consequences on career-relevant opportunities of affirmed individuals. In practice, the same performance evaluations are often used for multiple managerial decisions (Bracken 2001; Edwards and Ewen 1996; Johnson 2004). Distortions in performance evaluations arising due to a quota may therefore negatively affect the career opportunities of affirmed individuals in contexts which go beyond the scope of the quota regulation. Additionally, distorted peer-reviews may serve as an anchor for future performance evaluations (Tversky and Kahneman 1974; Thorsteinson et al. 2008). Furthermore, performance evaluations affect self-image and motivation (Heilman and Alcott 2001; Unzueta, Gutiérrez, and Ghavami 2010; Leslie, Mayer, and Kravitz 2014). Relatively unfavorable peer-reviews may therefore reduce the future productivity of

affirmed individuals and may discourage them from pursuing relevant career paths (Leibbrandt, Wang, and Foo 2018; Leibbrandt and List 2018).

In addition, we study spillover effects of quotas on altruistic behavior in a non-competitive context that is not regulated by a quota. In line with previous research on spillover effects of quotas (Kölle 2016; Banerjee, Gupta, and Villeval 2018; Maggiani and Montinari 2017), we do not find evidence of an effect of quotas on the overall level of altruistic behavior. However, the implementation of a quota in the previous experiment substantially reduces altruistic behavior among individuals who were affirmed in the previous experiment. Proponents of affirmative action argue that by increasing the representation of a minority group, this also fosters cooperative networks amongst them and thus acts as a catalyzer. Our findings, however, suggest that quota interventions may undermine the effectiveness with respect to the establishment of social networks (Athey, Avery, and Zemsky 1997; Buckles 2019; Bertrand et al. 2019).

Overall, our research provides an important contribution to better understand the mechanisms behind a quota intervention. While quotas are a popular form of affirmative action and have been rolled out in many countries, most of the literature has focused on the positive effects of quota interventions with respect to tournament entry and gender. Our paper enriches the discussion of quota interventions by exploring distortions in subjective peer-reviews as an important channel that could mitigate the positive effects of quotas.

## References

- 2020 Women on Boards. 2018. "2020 Women on Boards Gender Diversity Index: 2018 Progress of Women Corporate Directors by Company Size, State and Industry Sector."
- American Economic Association. 2018. "Report of the Committee on the Status of Minority Groups in the Economic Profession."
- Athey, Susan, Christopher Avery, and Peter Zemsky. 1997. "Mentoring and Diversity." *American Economic Review* 90 (4).
- Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017. "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability." *American Economic Review* 107 (3): 714–47.
- Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor. 2015. "Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India" 106 (6): 1495–1521.
- Balafoutas, Loukas, Brent J. Davis, and Matthias Sutter. 2016. "Affirmative Action or Just Discrimination? A Study on the Endogenous Emergence of Quotas." *Journal of Economic Behavior & Organization* 127: 87–98.
- Balafoutas, Loukas, and Matthias Sutter. 2012. "Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory." *Science* 335 (6068): 579–82.
- Balietti, Stefano, Robert L. Goldstone, and Dirk Helbing. 2016. "Peer Review and Competition in the Art Exhibition Game." *Proceedings of the National Academy of Sciences* 113 (30): 8414–19.
- Banerjee, Abhijit, Esther Duflo, Clement Imbert, and Rohini Pande. 2013. "Entry, Exit and Candidate Selection: Evidence from India." *Working Paper*.
- Banerjee, Ritwik, Nabanita Datta Gupta, and Marie Claire Villeval. 2018. "The Spillover Effects of Affirmative Action on Competitiveness and Unethical Behavior." *European Economic Review* 101: 567–604.
- Baskaran, Thushyanthan, and Zohal Hessami. 2018. "Does the Election of a Female Leader Clear the Way for More Women in Politics?" *American Economic Journal: Economic Policy* 10 (3): 95–121.
- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124 (4): 1497–1540.
- Beaurain, Guillaume, and David Masclet. 2016. "Does Affirmative Action Reduce Gender Discrimination and Enhance Efficiency? New Experimental Evidence." *European Economic Review* 90: 350–62.
- Beech, Nic, Nelarine Cornelius, Lisi Gordon, Geraldine Healy, Emmanuel Ogbonna, Gurchathen Sanghera, Chidozie Umeh, James Wallace, and Patrick Woodman. 2017. "Delivering Diversity."
- Bertrand, Marianne, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney. 2019. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway." *Review of Economic Studies* 86 (1): 191–239.
- Bertrand, Marianne, and Esther Duflo. 2016. "Field Experiments on Discrimination." 22014. NBER Working Paper. Cambridge.
- Bijkerk, Suzanne, Silvia Dominguez-Martinez, Jurjen Kamphorst, and Otto H. Swank. 2018. "Labor Market Quotas." TI 2018-036/VII. Tinbergen Institute Discussion Paper.
- Borgida, Eugene, and Susan T. Fiske. 2008. *Beyond Common Sense: Psychological Science in the Courtroom*. Malden, MA: Wiley-Blackwell.
- Bracken, David W. 2001. *The Handbook of Multisource Feedback: The Comprehensive Resource for Designing and Implementing MSF Processes*. San Francisco: Jossey-Bass.
- Buckles, Kasey. 2019. "Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage." *Journal for Economic Perspectives* 33 (1): 43–60.
- Calsamiglia, Caterina, Jörg Franke, and Pedro Rey-Biel. 2013. "The Incentive Effects of Affirmative Action in a Real-Effort Tournament." *Journal of Public Economics* 98: 15–31.
- Cappelen, Alexander W., Astri Drange Hole, Erik Sorensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3): 818–27.
- Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm. 2010. "Tournaments and Office Politics:



- Evidence from a Real Effort Experiment.” *The American Economic Review* 100 (1): 504–17.
- Chowdhury, Subhasish M., and Oliver Gürtler. 2015. “Sabotage in Contests: A Survey.” *Public Choice* 164 (1–2): 135–55.
- Coffman, Katherine B. 2014. “Evidence on Self-Stereotyping and the Contribution of Ideas.” *The Quarterly Journal of Economics* 129 (4): 1625–1660.
- Croson, Rachel, and Uri Gneezy. 2009. “Gender Differences in Preferences.” *Journal of Economic Literature* 47 (2): 448–74.
- Edwards, Mark R., and Ann J. Ewen. 1996. “How to Manage Performance and Pay with 360-Degree Feedback.” *Compensation & Benefits Review* 28 (3): 41.
- European Commission. 2016. “Gender Balance on Corporate Boards: Europe Is Cracking the Glass Ceiling.”
- Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton. 1994. “Fairness in Simple Bargaining Experiments.” *Games and Economic Behavior* 6 (3): 347–369.
- Franke, Jörg. 2012. “Affirmative Action in Contest Games.” *European Journal of Political Economy* 28 (1): 105–18.
- Greenberg, Jerald. 1990. “Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts.” *Journal of Applied Psychology* 75 (5): 561–68.
- Greiner, Ben. 2015. “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE.” *Journal of the Economic Science Association* 1 (1): 114–125.
- Gürtler, Marc, and Oliver Gürtler. 2015. “Promotion Signaling, Discrimination, and Positive Discrimination Policies.” No. D04-V2. Beiträge Zur Jahrestagung Des Vereins Für Socialpolitik 2015: Ökonomische Entwicklung - Theorie Und Politik - Session: Labor - Theory.
- Harbring, Christine, Bernd Irlenbusch, Matthias Kräkel, and Reinhard Selten. 2007. “Sabotage in Asymmetric Contests - An Experimental Analysis.” *International Journal of the Economics of Business* 14: 201–23.
- Harrison, David A., David A. Kravitz, David M. Mayer, Lisa M. Leslie, and Dalit Lev-Arey. 2006. “Understanding Attitudes toward Affirmative Action Programs in Employment: Summary and Meta-Analysis of 35 Years of Research.” *Journal of Applied Psychology* 91 (5): 1013–36.
- Heilman, Madeline E., and Victoria Barocas Alcott. 2001. “What I Think You Think of Me: Women’s Reaction to Being Viewed as Beneficiaries of Preferential Selection.” *Journal of Applied Psychology* 86 (4): 574–82.
- Heilman, Madeline E., Caryn J. Block, and Jonathan A. Lucas. 1992. “Presumed Incompetent? Stigmatization and Affirmative Action Efforts.” *Journal of Applied Psychology* 77 (4): 536–44.
- Heilman, Madeline E., Caryn J. Block, and Peter Stathatos. 1997. “The Affirmative Action Stigma of Incompetence: Effects of Performance Information Ambiguity.” *The Academy of Management Journal* 40 (3): 603–25.
- Holzer, Harry, and David Neumark. 2000. “Assessing Affirmative Action.” *Journal of Economic Literature* 38 (3): 483–568.
- Houser, Daniel, Stefan Vetter, and Joachim Winter. 2012. “Fairness and Cheating.” *European Economic Review* 56 (8): 1645–55.
- Ip, Edwin, Andreas Leibbrandt, and Joseph Vecchi. forthcoming. “How Do Gender Quotas Affect Workplace Relationships? Complementary Evidence from a Representative Survey and Labor Market Experiments.” *Management Science*.
- Johnson, Lauren. 2004. “The Ratings Game: Retooling 360s for Better Performance.” *Harvard Management Update* 1: 1–4.
- Kölle, Felix. 2016. “Affirmative Action and Team Performance.” 2016–07. CeDEx Discussion Paper Series.
- Konow, James. 1996. “A Positive Theory of Economic Fairness.” *Journal of Economic Behavior and Organization* 31 (1): 13–35.
- Laske, Katharina, and Marina Schröder. 2016. “Quantity, Quality, and Originality: The Effects of Incentives on Creativity.” 7–1. CGS Working Paper.
- Leibbrandt, Andreas, and John A. List. 2018. “Do Equal Opportunity Statements Backfire? Evidence From a Natural Field Experiment on Job-Entry Decisions.” 25035. NBER Working Paper Series.
- Leibbrandt, Andreas, Liang Choon Wang, and Cordelia Foo. 2018. “Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash Against Women.” *Management Science* 64 (8): 3501–16.

- Leiner, D. J. 2014. "SoSci Survey (Version 2.5.00-i) [Computer Software]." 2014.
- Leslie, Lisa M., David M. Mayer, and David A. Kravitz. 2014. "The Stigma of Affirmative Action: A Stereotyping-based Theory and Meta-Analytic Test of the Consequences for Performance." *Academy of Management Journal* 57 (4): 964–89.
- Lundberg, Shelly, and Jenna Stearns. 2019. "Women in Economics: Stalled Progress." *Journal for Economic Perspectives* 33 (1): 3–22.
- Maggian, Valeria, and Natalia Montinari. 2017. "The Spillover Effects of Gender Quotas on Dishonesty." *Economics Letters* 159: 33–36.
- Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2019. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association* 17 (2): 535–566.
- 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.
- Niederle, Muriel, and Lise Versterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122 (3): 1067–1101.
- Nieva, Veronica F., and Barbara A. Gutek. 1980. "Sex Effects on Evaluation." *Academy of Management Review* 5 (2): 267–76.
- Rosaz, Julie, and Marie Claire Villeval. 2012. "Lies and Biased Evaluation: A Real-Effort Experiment." *Journal of Economic Behavior and Organization* 84 (2): 537–49.
- Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." *American Economic Review: Papers & Proceedings* 107 (5): 141–45.
- Schotter, Andrew, and Keith Weigelt. 1992. "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results." *The Quarterly Journal of Economics* 107 (2): 551–539.
- Schweitzer, Maurice E., and Donald E. Gibson. 2008. "Fairness, Feelings, and Ethical Decision-Making: Consequences of Violating Community Standards of Fairness." *Journal of Business Ethics* 77: 287–301.
- Selten, Reinhard. 1998. "Axiomatic Characterization of the Quadratic Scoring Rule." *Experimental Economics* 1: 43–62.
- Sowell, Thomas. 2004. *Affirmative Action around the World: An Empirical Study*. New Haven & London: Yale University Press.
- Thorsteinson, Todd J., Jennifer Breier, Anna Atwell, Catherine Hamilton, and Monica Privette. 2008. "Anchoring Effects on Performance Judgments." *Organizational Behavior and Human Decision Processes* 107 (1): 29–40.
- Turner, Marlene F., and Anthony R. Pratkanis. 1993. "Effects of Preferential and Meritorious Selection on Performance: An Examination of Intuitive and Self-Handicapping Perspectives." *Personality and Social Psychology Bulletin* 19 (1): 47–58.
- Tversky, A., and D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.
- Unzueta, Miguel M., Angélica S. Gutiérrez, and Negin Ghavami. 2010. "How Believing in Affirmative Action Quotas Affects White Women's Self-Image." *Journal of Experimental Social Psychology* 46 (1): 120–26.
- Wallon, Gerlind, Sandra Bendiscioli, and Michelle S. Garfinkel. 2015. "Exploring Quotas in Academia."

## **For Online Publication**

### **Appendix 1: Experimental Instructions – Main Experiment**

Welcome to this experiment!

Please carefully read through the following instructions. If you have a question, please raise your hand. We will then come to your desk and answer your question.

All of your decisions are anonymous. Communication with other participants is not permitted for the duration of the experiment. We would like to ask you to switch off your mobile phone and place it in your bag.

You will receive a show-up fee of 4 euros for your participation. You can earn additional money in the following experiment.

## Instructions

### Experimental Procedure

- This experiment consists of **multiple rounds**.
  - Initially, you will take part in a test round that is not relevant for your payment.
  - After that, 3 experimental rounds will be conducted.
- At the beginning of the experiment, you will be randomly assigned a type.
  - You are either **type “green”** or **type “yellow”**.
  - You can recognize your type based on the color of the frame of your display.
  - This type assignment remains constant for the entire experiment.
- In each round, you will be randomly assigned to a group of four participants.
  - Each group consists of 2 type “green” and 2 type “yellow” participants.
  - You will be assigned to a new group in each round.
  - However, your type (green or yellow) remains constant for the entire experiment.

### Your Task

- Your task is to **illustrate an object using given materials**.
  - Group members of both types are provided with the same materials to illustrate the object (see images below)
    - The materials provided to type “yellow” members are pictured in the bottom left (yellow frame).
    - The materials provided to type “green” members are pictured in the bottom right (green frame).
  - The object that you are supposed to illustrate will be displayed on your screen.
  - You will illustrate a different object in each round.
  - All members of your group have to illustrate the same object in the respective rounds.

#### *Baseline\_equal and Quota\_equal Treatments:*

- You have **5 minutes** time available in each round.

#### *Baseline\_unequal and Quota\_unequal Treatments:*

- The time available to illustrate the object in each round is limited
  - **Type “yellow” (“green”)** group members have 2 minutes and 30 seconds available in each round.
  - **Type “green” (“yellow”)** group members have 5 minutes available in each round.



### Rating the Illustrations

- Each group member rates the other 3 group members' illustrations
- The color of the frame of the respective illustrations indicates the type (green or yellow) to which the group member who produced the illustration belongs.
- The rating is conducted on a scale from **0.0 to 10.0 points**.
  - 0.0 points correspond to the worst rating.
  - 10.0 points correspond to the best rating.
  - Please always specify **exactly one decimal place** (please use a dot as a decimal sign).
- The sum of the points awarded equals the final rating.
  - For each illustration, this **final rating** lies **between 0.0 and 30.0 points**.

### Payment

- At the end of the experiment, **one of the three experimental rounds** will be randomly **chosen**.
  - Only this round is relevant for the payment.
- Based on the final rating and the assigned type, exactly **two prizes** with the amount of **16 euros** each will be paid out to **two** different group members.

### Baseline\_equal and Baseline\_unequal Treatments:

- The group member with the highest final rating **among all** group members of **both types** receives a prize.
- The group member with the highest final rating among the **remaining three** group members of **both types** receives a prize.
- The other two group members do not receive a prize.
- This means that **at most two prizes** are awarded to **type "green"** group members.

- This means that **at most two prizes** are awarded to **type “yellow”** group members.

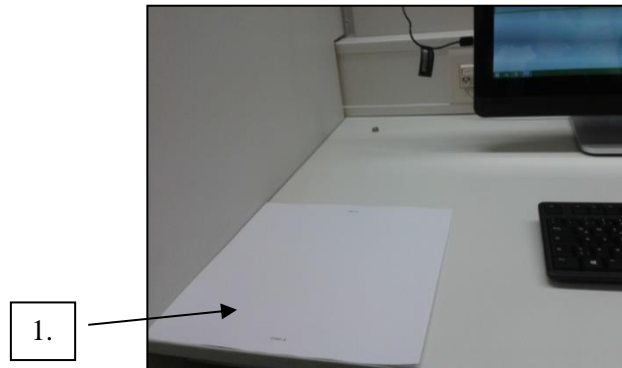
*Quota\_equal and Quota\_unequal Treatments:*

- The **type “green” (“yellow”)** group member with the higher final rating between **the two** type “green” (“yellow”) group members receives a prize.
- The group member with the highest final rating among the **remaining three** group members of **both types** receives a prize.
- The other two group members do not receive a prize.
- This means that **at least one prize and at most two prizes** are awarded to **type “green” (“yellow”)** group members.
- This means that **at most one prize** are awarded to the **type “yellow” (“green”)** group members.

## Procedure

Please proceed with the illustration of the object as follows:

1. Illustrate the object in the marked area using the provided materials.



2. Take a photo of the illustrated object by clicking on “take photo.” If the photo meets your expectations, save this photo by clicking on “save photo.” If a photo does not meet your expectations, you can delete it by clicking on “delete photo.”
3. You can take and save as many photos as you wish during the time available. You cannot take or save any additional photos after time has expired.
4. After time has expired, you have the opportunity to choose one of your saved photos. Only this photo will be rated by the other group members in the next step. None of the other photos will be rated.
5. Note that only previously saved photos can be chosen.

Please note the following when illustrating the objects:

- **Only** use the provided materials.
- For each illustration, you may use all materials or a selection of materials.
- Place the object only **within** the area marked with the piece of paper (only this area will be captured by the camera)
- Pay attention to the **direction** of your illustration (the piece of paper is labeled with “top” and “bottom”).
- Pay attention that **your hands** are **not visible** in the marked area.
- Keep the **unused materials outside** of the marked area.
- Please **do not write or draw** on the piece of paper representing the marked area.
- Pay attention to the **time limit** for the task; after time has expired you **cannot** take or save any **new photos**.

## Appendix 2: Questionnaire

### Questionnaire

Please, answer the following questions while we prepare the payment.  
Thank you!

1. Please, indicate your gender:	female <input type="checkbox"/>	male <input type="checkbox"/>
2. How old are you?		
3. How many siblings do you have?		
4. Do you have a red-green colorblindness?	Yes <input type="checkbox"/>	No <input type="checkbox"/>
<b>Please answer the following questions using the provided scale.</b>		
5. How satisfied are you with the experiment overall?		
Not satisfied at all <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Very satisfied <input type="checkbox"/>		
6. How much would you like to participate in an experiment like this one again?		
Not at all <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Very much <input type="checkbox"/>		
7. How likely is it that you would recommend participating in an experiment like this one to a friend?		
Not likely at all <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Very likely <input type="checkbox"/>		



8. How fair did you consider the payment procedure of this experiment?				
Not fair at all				Very fair
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Did you feel disadvantaged or advantaged by the payment procedure of this experiment?				
Very disadvantaged				Very advantaged
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. How justified did you consider this disadvantage or advantage?				
Not justified at all				Very justified
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. How much fun did you have solving the task?				
No fun at all				A lot of fun
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. How creative are you?				
Not creative at all				Very creative
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. How difficult did you find the task?				
Not difficult at all				Very difficult
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

14. How well, do you think, you performed compared to the other participants in this room?				
Far below average				Far above average
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. In how many rounds do you think you received the highest or second highest final rating?				
0	1	2	3	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
16. In how many of these three rounds do you think you received a prize?				
0	1	2	3	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
17. How likely is it, do you think, that the other group members rated your illustration according to its quality?				
Not likely at all				Very likely
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. How do you see yourself: Are you someone who is willing to take risks or do you try to avoid them?				
Not willing at all to take risks				Very willing to take risks
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<b>Please indicate for the following questions how much the respective statements apply to you.</b>				
19. I get upset when someone is better off for no reason.				
Doesn't apply at all				Fully applies
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. I feel guilty when I am better off for no reason.				
Doesn't apply at all				Fully applies
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. If someone harms me on purpose, I will try to take revenge even when I have to bear the costs.				
Doesn't apply at all				Fully applies
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. If someone does me a favor, I am happy to return it.				
Doesn't apply at all				Fully applies
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. I like to compete with others.				
Doesn't apply at all				Fully applies
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. It is important to me to be the best.				
Doesn't apply at all				Fully applies
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Appendix 3: Summary Statistics

**Table A1: Summary statistics**

	<i>Baseline_equal</i>	<i>Quota_equal</i>		<i>Baseline_unequal</i>		<i>Quota_unequal</i>	
		Affirmed	Non-affirmed	Disadvantaged	Non-disadvantaged	Affirmed	Non-affirmed
Peer-review received	2.88 (0.77)	2.27 (0.70)	2.83 (0.67)	2.37 (0.92)	2.38 (1.00)	2.37 (0.89)	3.02 (0.78)
Independent rating received	6.28 (0.38)	5.98 (0.62)	6.16 (0.56)	5.71 (0.70)	6.21 (0.59)	5.46 (0.69)	6.18 (0.74)
Independent rating received (lab rater)	4.54 (0.47)	4.51 (0.77)	4.51 (0.74)	n.a.	n.a.	n.a.	n.a.
Procedural fairness perception	2.39 (0.40)	2.26 (0.42)	1.81 (0.50)	2.11 (0.48)	2.42 (0.54)	2.55 (0.54)	2.24 (0.53)
Female	0.56 (0.20)	0.66 (0.30)	0.59 (0.23)	0.48 (0.27)	0.63 (0.22)	0.63 (0.23)	0.65 (0.23)

Means with standard errors on group level in parentheses. Evaluation scale for procedural fairness perception: 1 – very unfair to 5 – very fair

## Appendix 4: Regression Analysis Peer-reviews Provided by Rater Type with Lab Raters as Control for Independent Ratings

**Table A2: Regression analysis peer-reviews provided by rater type with lab raters as control for independent ratings**

<i>Dependent variable:</i> Peer-review	(1) All	Ex-ante equal (2) Affirmed rater	(3) Non-affirmed rater
Quota	-0.210 (0.224)	0.036 (0.278)	-0.876*** (0.313)
Affirmed ratee	-0.565*** (0.145)	-0.659*** (0.194)	0.075 (0.085)
Independent rating (lab rater)	0.427*** (0.044)	0.470*** (0.053)	0.421*** (0.048)
Constant	0.940*** (0.232)	0.749*** (0.262)	0.968*** (0.241)
Observations	2,880	2,160	2,160
Number of participants	320	240	240
Number of groups	40	40	40

Two-way error component linear model, allowing for creator and evaluator random effects. Separate models for all or affirmed and non-affirmed raters. The dependent variable is peer-review received. Independent variables: Quota (dummy equal to one in treatments with a quota), Affirmed (dummy equal to one for affirmed ratee in the treatments with a quota), Independent rating (lab rater) (continuous variable with the mean evaluation of independent lab raters who are blind to treatments but have previous experience with the task). All individuals from the *Baseline\_equal* serve as a reference group for both affirmed and non-affirmed raters.

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

## Appendix 5: Regression Analysis Peer-reviews Provided - All Treatments Combined

**Table A3: Regression analysis peer-reviews provided – all treatments combined**

<i>Dependent variable: Peer-review</i>	(1) All	(2) Affirmed rater	(3) Non-affirmed rater
<i>Quota_equal</i>	-0.012 (0.241)	0.389 (0.299)	-0.801*** (0.281)
<i>Quota_equal</i> x affirmed ratee	-0.498*** (0.142)	-0.595*** (0.188)	0.142 (0.089)
Unequal	-0.470 (0.296)	-0.602* (0.314)	-0.193 (0.319)
Unequal x disadvantaged ratee	0.169 (0.117)	0.003 (0.116)	0.055 (0.142)
<i>Quota_unequal</i>	0.650** (0.296)	1.028*** (0.342)	-0.099 (0.402)
<i>Quota_unequal</i> x affirmed ratee	-0.549*** (0.198)	-0.701*** (0.196)	0.103 (0.177)
Independent Rating	0.362*** (0.018)	0.347*** (0.026)	0.365*** (0.022)
Constant	0.603*** (0.195)	0.700*** (0.218)	0.580*** (0.200)
Observations	5,688	3,564	3,564
Number of participants	632	316	316
Number of groups	79	79	79

Two-way error component linear model, allowing for creator and evaluator random effects. Models including all settings and treatments and either all or only affirmed or non-affirmed raters. The dependent variable is peer-review received. Independent variables: *Quota\_equal* (dummy equal to one in *Quota\_equal* treatment), *Quota\_equal* x Affirmed ratee (dummy equal to one for affirmed ratee in the *Quota\_equal* treatment), Unequal (dummy equal to one for ex-ante unequal treatment), Unequal x Disadvantaged ratee (dummy equal to one for disadvantaged ratee in ex-ante unequal treatment), *Quota\_unequal* (dummy equal to one in *Quota\_unequal* treatment), *Quota\_unequal* x Affirmed ratee (dummy equal to one for affirmed ratee in the *Quota\_unequal* treatment), Independent Rating (continuous variable with the mean evaluation of independent raters who are blind to treatments).

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

## Appendix 6: Regression Analysis Independent Ratings

**Table A4: Regression analysis independent ratings**

<i>Dependent variable: Independent rating</i>	(1) Ex-ante equal	(2) Ex-ante unequal
Quota	-0.125 (0.150)	-0.029 (0.211)
Affirmed ratee	-0.177 (0.149)	-0.229 (0.238)
Disadvantaged ratee		-0.495*** (0.183)
Constant	6.281*** (0.084)	6.209*** (0.130)
Observations	960	936
Number of participants	320	312
Number of groups	79	79

Ordinary least squares linear model. Models include either ex-ante equal or ex-ante unequal settings. The dependent variable is independent rating received. Independent variables: Quota (dummy equal to one in treatments with a quota), Affirmed ratee (dummy equal to one for affirmed ratee in the treatments with a quota), Disadvantaged ratee (dummy equal to one for disadvantaged ratee in all treatments involving less working time for disadvantaged type).

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## Appendix 7: Regression Analysis Peer-reviews Provided by Rater Type with Control for Perceived Procedural Fairness

**Table A5: Regression analysis peer-reviews provided by rater type with control for perceived procedural fairness**

<i>Dependent variable:</i> Peer-review	Ex-ante equal		Ex-ante unequal	
	(1) Affirmed rater	(2) Non-affirmed rater	(3) Affirmed rater	(4) Non-affirmed rater
Quota	0.335 (0.282)	-0.315 (0.275)	0.619* (0.353)	-0.009 (0.400)
Affirmed ratee	-0.595*** (0.190)	0.137 (0.088)	-0.702*** (0.199)	0.112 (0.178)
Disadvantaged ratee			0.000 (0.117)	0.073 (0.144)
Independent rating	0.346*** (0.031)	0.342*** (0.025)	0.342*** (0.042)	0.401*** (0.035)
Fairness perception of rater	0.838*** (0.140)	0.861*** (0.138)	0.931*** (0.235)	0.485*** (0.150)
Constant	-1.291*** (0.380)	-1.327*** (0.392)	-1.835*** (0.529)	-0.981** (0.436)
Observations	2,160	2,160	1,404	1,404
Number of participants	320	320	312	312
Number of groups	40	40	39	39

Two-way error component linear model, allowing for creator and evaluator random effects. Separate models for affirmed and non-affirmed raters. The dependent variable is peer-review received. Independent variables: Quota (dummy equal to one in treatments with a quota), Affirmed ratee (dummy equal to one for affirmed ratee in the treatments with a quota), Disadvantaged ratee (dummy equal to one for disadvantaged ratee in all treatments involving less working time for disadvantaged type), Independent Rating (continuous variable with the mean evaluation of independent raters who are blind to treatments), Fairness perception of rater (continuous variable with the rater's evaluation of fairness as elicited in the post-experimental questionnaire). In the ex-ante equal setting, all individuals from the *Baseline\_equal* serve as a reference group for both affirmed and non-affirmed raters. In the ex-ante unequal, the disadvantaged individuals from the *Baseline\_unequal* treatment serve as the reference group for the affirmed types, while the non-disadvantaged individuals from the *Baseline\_unequal* treatment serve as reference group for non-affirmed types in the *Quota\_unequal* treatment.

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.



## Appendix 8: Regression Analysis Peer-reviews Provided by Rater Type – Gender Differences

**Table A6: Regression analysis peer-reviews provided by rater type – gender differences**

Dependent variable: Peer-review	Ex-ante equal				Ex-ante unequal			
	Affirmed rater		Non-affirmed rater		Affirmed rater		Non-affirmed rater	
	(1) Female	(2) Male	(3) Female	(4) Male	(5) Female	(6) Male	(7) Female	(8) Male
Quota	0.976** (0.409)	-0.259 (0.472)	-0.711* (0.405)	-0.777* (0.415)	2.052*** (0.542)	-0.659 (0.568)	0.148 (0.486)	-0.704 (0.724)
Affirmed ratee	-0.562*** (0.193)	-0.654* (0.338)	0.113 (0.123)	0.183* (0.103)	-0.708*** (0.226)	-0.638 (0.406)	0.118 (0.220)	0.091 (0.311)
Disadvantaged ratee					-0.021 (0.126)	0.034 (0.227)	0.062 (0.176)	0.032 (0.191)
Independent rating	0.395*** (0.041)	0.310*** (0.053)	0.367*** (0.038)	0.316*** (0.047)	0.411*** (0.071)	0.317*** (0.055)	0.429*** (0.049)	0.314*** (0.058)
Constant	0.508 (0.350)	0.355 (0.331)	0.690** (0.328)	0.318 (0.303)	-0.445 (0.536)	0.604* (0.351)	0.095 (0.381)	0.189 (0.579)
Observations	1,278	882	1,224	936	774	630	891	513
Number of participants	142	98	136	104	86	70	99	57
Number of groups	39	35	40	39	37	35	39	33

Two-way error component linear model, allowing for creator and evaluator random effects. Separate models for affirmed and non-affirmed and female and male raters. The dependent variable is peer-review received. Independent variables: Quota (dummy equal to one in treatments with a quota), Affirmed ratee (dummy equal to one for affirmed ratee in the treatments with a quota), Disadvantaged ratee (dummy equal to one for disadvantaged ratee in all treatments involving less working time for disadvantaged type), Independent Rating (continuous variable with the mean evaluation of independent raters who are blind to treatments). In the ex-ante equal setting, all individuals from the *Baseline\_equal* serve as a reference group for both affirmed and non-affirmed raters. In the ex-ante unequal, the disadvantaged individuals from the *Baseline\_unequal* treatment serve as the reference group for the affirmed types, while the non-disadvantaged individuals from the *Baseline\_unequal* treatment serve as reference group for non-affirmed types in the *Quota\_unequal* treatment.

In all specifications robust standard errors are clustered by matching groups of eight participants. Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

## Appendix 9: Experimental Instructions - Dictator Experiment

### Instructions

Welcome to this part of the experiment! Please carefully read through the following instructions. If you have questions, please raise your hand. We will then come over to you and answer your question. As before, all of your decisions are anonymous. Communication with other participants is not permitted for the duration of the experiment.

### The Experiment

This experiment consists of four rounds. In each round you will be **anonymously assigned** to another participant and also **assigned** a role. At the beginning of each round, you will learn the other participant's type (green or yellow) as well as your role (active or passive) in this round. Assignment to the player types (green or yellow) corresponds with the assignment from the first experiment. **The roles will be newly assigned in each round.**

- The **active** participant has an **endowment of 2.00 euros** in each round.
- The **passive** participant has an **endowment of 0.00 euro** in each round.
- The **active** participant **chooses** how many euros (in 0.10 euro increments) he/she would like to **transfer to** the **passive** participant.

A total of **4 rounds** will be conducted in this experiment. The assignment of roles (active and passive) varies between each round, and you will be assigned to a different participant in each round.

At the end of the experiment, **one round** will be **randomly** determined to be relevant for the payment.

### Payment

- The active participant's payment equals 2.00 euros minus the amount transferred in the round relevant for the payment, i.e., **the active participant's payment = 2.00 euros – transferred amount.**
- The passive participant's payment equals 0.00 euro plus the amount transferred in the round relevant for the payment, i.e., **the passive participant's payment = 0.00 euro + transferred amount.**