IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13202

Perceived Fairness and Consequences of Affirmative Action Policies

Hannah Schildberg-Hörisch
Chi Trieu
Jana Willrodt

APRIL 2020

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13202

Perceived Fairness and Consequences of Affirmative Action Policies

Hannah Schildberg-Hörisch
DICE and IZA

Chi Trieu
DICE

Jana Willrodt
DICE

APRIL 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany
Phone: +49-228-3894-0
Email: publications@iza.org
www.iza.org

ABSTRACT

Perceived Fairness and Consequences of Affirmative Action Policies*

Debates about affirmative action often revolve around fairness. Accordingly, we document substantial heterogeneity in the fairness perception of various affirmative action policies. But do these differences translate into different consequences? In a laboratory experiment, we study three different quota rules that favor individuals whose performance is low, either due to bad luck (discrimination), low productivity, or choice of a short working time. Higher fairness perceptions coincide with a higher willingness to compete and less retaliation against winners. No policy harms overall efficiency or post-competition teamwork. Furthermore, individuals seem to internalize the norm behind the policies that are perceived as fairest.

JEL Classification: C91, D02, D63

Keywords: affirmative action, fairness ideals, experiment, tournament, real effort

Corresponding author:
Hannah Schildberg-Hörisch
Düsseldorf Institute for Competition Economics (DICE)
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
40225 Düsseldorf
Germany

E-mail: schildberg-hoerisch@dice.hhu.de

* We thank Björn Bartling, Lorenz Götte, Paul Heidhues, Muriel Niederle, Hans-Theo Normann, Jan Potters, Gerhard Riener, Bettina Rockenbach, Daniel Schunk, Matthias Sutter, Marie Claire Villeval and Roberto Weber for useful comments and suggestions. Additionally, we thank participants of ESA 2018 in Berlin, EWEBE 2018 in Tilburg, MBEPS 2018 in Maastricht and seminars in Aachen, Cologne, Frankfurt and Göttingen, at MPI for Research on Collective Goods in Bonn and at DICE for helpful comments and discussions. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through 235577387/GRK1974 is gratefully acknowledged.

1 Introduction

Ever since its introduction, affirmative action has been subject to heated debates (see, e.g., Fish, 2000; Fullinwider, 2011).[1] A common argument professed by its opponents is that under affirmative action decisions such as university admission, hiring or promotion choices are not purely based on merit. This goes against the ideal of a "fair" policy that should select or reward the best performers. Proponents of affirmative action on the other hand may argue that a truly fair policy should take disadvantages (e.g., due to family background, poverty, race, and gender) into account when evaluating performance to compensate for a lack of opportunities due to discrimination, historical injustice, or the "accident of birth".[2]

The underlying conflict in this debate is the collision of different fairness ideals originating from different theories of distributive justice. Each professes a distinct view on which factors of their performance individuals should be held accountable for. The three stylized factors that determine an individual's performance are effort, ability, and luck (Cappelen et al., 2007, 2010).

From a *libertarian* point of view, individuals should be held accountable for all three factors, making affirmative action unnecessary. *Meritocracism*, however, posits that only personal factors, i.e., effort and ability, should be considered when assessing someone's performance, justifying affirmative action that compensates for differences in luck. According to *choice egalitarianism*, people should be held accountable only for factors within their control. This means affirmative action policies should offset bad luck and differences in ability, but not self-chosen effort. Finally, *strict egalitarianism* strives for complete distributive equality, thus calling for policies that counterbalance differences in all three factors.

Despite a growing amount of empirical evidence on the consequences of affirmative action, the perceived fairness of different affirmative action policies and the implications of those fairness perceptions for the effectiveness of affirmative action policies have barely been investigated. In this paper, we provide evidence from a laboratory experiment that implements several affirmative action policies in the form of quota rules in a tournament setting, and explicitly links a broad set of outcomes of those policies to their fairness perception. As a further novelty, the affirmative action policies address all three determinants of performance in a unified framework. One policy reflects meritocracism, favoring individuals disadvantaged due to persistent bad luck, an impersonal factor that resembles discrimination.[3] Another favors those disadvantaged by low innate ability (a personal, but out-of-

[1] Affirmative action is defined as a policy that promotes the opportunities of defined minority groups within a society to give them equal access to that of the privileged majority population (Affirmative Action, Harvard Law School Blog, accessed April 27, 2020).

[2] Another argument in favor of affirmative action policies with regard to, e.g., university admission, hiring or promotions, is the need for a more diverse student body or employee composition further up the hierarchy. This aspect is beyond the scope of this paper.

[3] In our design, (bad) luck is a randomly assigned characteristic and remains unchanged for each subject throughout the experiment. This characteristic resembles stable causes of discrimination such as gender, skin color, or race. Similarly, Balafoutas et al. (2016), Calsamiglia et al. (2013), Fal-

control factor) proxied by productivity in the experimental task in which practicing does not improve performance. A third policy favors individuals whose performance is lower due to providing less effort at the extensive margin by choosing a shorter working time (an in-control factor). To our knowledge, we are the first to investigate affirmative action policies that compensate for differences in working time or productivity – although similar policies exist outside the laboratory.[4]

In particular, we address three related research questions. Do the consequences of affirmative action policies depend on which of the three factors (bad luck, low productivity, or short working time) they compensate for? Are the three affirmative action policies perceived as differently fair? And is there a link between the perceived fairness of these policies and their consequences? Answering these questions is key to understanding the approval and the implications of different affirmative action policies.

To address these questions, we elicit individual fairness perceptions for all affirmative action schemes and consider a broad set of outcomes. First, we look at the immediate consequences of affirmative action within the tournament, namely willingness to compete and efficiency (in terms of output produced). We then explore potential spillover effects on post-competition outcomes when affirmative action is no longer in place. For that purpose, we elicit separate measures for cooperation in a team and spiteful behavior targeting those favored by affirmative action ("retaliation").

We find that none of the affirmative action policies is considered less fair than no affirmative action, documenting wide acceptance for affirmative action from a normative fairness point of view. Still, heterogeneity in the perceived fairness of the different affirmative action policies is substantial: affirmative action targeting bad luck (discriminated) individuals is perceived as fairest, followed by a policy in favor of individuals choosing a short working time, while affirmative action targeting individuals with low productivity and no affirmative action are perceived as equally and least fair.

Importantly, none of the three affirmative action policies under study harms overall willingness to compete or efficiency. However, the heterogeneity in fairness perceptions goes hand in hand with more detailed consequences of these policies. Targeted subjects are more likely to compete under a quota rule instead of working under a piece rate payment the fairer the quota rule is *generally* perceived to be. Additionally, non-targeted subjects are more willing to enter a tournament with a quota rule if they *personally* perceive the quota rule as fair.

---

lucchi and Quercia (2018), and Petters and Schroeder (2019) analyze affirmative action policies that compensate for differences in a randomly assigned, exogenously given characteristic.

[4] For example, the Council Directive 97/81/EC of the European Union states that part-time employees may not be treated less favorably than full-time employees (see Council of European Union, 1997). This includes their equal access to promotions (although their overall performance in terms of output is typically lower). Other policies are designed to compensate for worse performance due to low productivity. For example, students with dyslexia or physical restraints such as typist's cramps or poor eyesight can get some extra time in exams (see, e.g., Disability Rights Commission, 2007).

Regarding potential spillover effects we observe no difference between policies in post-competition teamwork. Also, no policy induces retaliation at the aggregate level. Finally, we provide first evidence on internalization of those affirmative action policies that are rated as fairest: subjects still support individuals with bad luck and short working time in post-competition interactions in which the corresponding affirmative action policies are no longer in place.

In sum, the fairness perception of an affirmative action policy seems to shape its consequences. This is an important insight for the successful implementation of such policies if they are politically desired. For example, providing evidence on existing discrimination against a targeted group may help to increase the acceptance of an affirmative action policy and, in turn, positively impact its consequences.

Our paper contributes to a growing literature researching the effectiveness of affirmative action. Schotter and Weigelt (1992) and Calsamiglia et al. (2013) show that affirmative action in the form of bonuses or lump-sum payments for subjects who face an exogenous disadvantage in competitions can increase their performance. Many studies consider gender quotas or other preferential treatment of women in labor-market related settings and find positive overall effects (e.g., Beaman et al., 2009; Balafoutas and Sutter, 2012; Beaurain and Masclet, 2016; Niederle et al., 2013). For example, there is evidence that such policies increase women's willingness to enter competitions without discouraging men (Balafoutas and Sutter, 2012; Niederle et al., 2013; Ibanez and Riener, 2018). This is also true when gender quotas are introduced endogenously by vote (Balafoutas et al., 2016). Kölle (2017) finds that gender quotas neither harm effort provision within teams nor the willingness to work in teams. Beside gender, several studies investigate affirmative action for members of disadvantaged castes in India (Banerjee et al., 2018, 2019; Bagde et al., 2016; Jensenius, 2015). For example, Banerjee et al. (2018) show that affirmative action boosts confidence and willingness to compete of targeted subjects, but this effect disappears when affirmative action is removed.

However, there is also evidence of adverse consequences of affirmative action (Fallucchi and Quercia, 2018; Heilman et al., 1997; Leibbrandt et al., 2017; Leibbrandt and List, 2018). In particular, in Leibbrandt et al. (2017) a gender quota turns women into the target of sabotage, thereby undermining their willingness to compete. Similarly, Fallucchi and Quercia (2018) find that the threat of retaliation reduces competition entry of targeted subjects.

These seemingly contradictory findings bring up the question under which conditions affirmative action has adverse consequences. Answering it may provide valuable insights into how affirmative action policies that are politically desired can be implemented without causing more harm than good. The studies of Ip et al. (2020) and Petters and Schroeder (2019) provide first related evidence. In a gift-exchange game with payoffs that depend on manager productivity, Ip et al. (2020) find that quotas for female managers decrease workers' effort when women are perceived as having lower skills than men, but not when they are discriminated against in the manager selection process. In an independent, representative survey with US citizens, approval for gender quotas for leadership positions is high when women are discriminated against in the recruitment process, but low otherwise (regardless of whether a gender skill gap exists). Petters and Schroeder (2019) study the effect of randomly assigned quotas on peer-ratings of performance and find that targeted individuals' performance is rated worse than that of non-targeted individuals with a similar performance.

These studies indicate that the effects of affirmative action policies can depend crucially on whether and how they are justified, which in turn may impact their perception as more or less fair. This observation lays the ground for jointly studying perceived fairness and effects on outcomes of such policies. Inspired by real-world policies, criteria for affirmative action in most existing laboratory and field experiments are gender or ethnicity. However, such policies are sure to be perceived differently by different people. Take the most widely studied example of a gender quota: if a woman's performance is not among the best, some might perceive this as being the result of discrimination, while others may attribute her performance to low innate productivity, or a personal choice of working part-time. Usually, we cannot observe which of the three perceptions (or a mixture thereof) is invoked, although this is crucial to understanding the reaction to a gender quota.

To avoid this problem, we study affirmative action in a more stylized environment. In particular, we investigate the results of three different policies, each based on one of the three separate determinants of performance: persistent bad luck (resembling discrimination), effort (measured by self-chosen working time), and innate productivity. Compared to quotas for women or minorities, these affirmative action policies explicitly state the reason for a favored treatment of the respective target group. As a consequence, different judgments regarding their fairness can be unequivocally attributed to holding different fairness ideals instead of possibly reflecting different perceptions of the reasons for the target group's favored treatment. Our design thus provides a sound basis for analyzing whether and how the fairness perception of a specific affirmative action policy impacts its consequences.

Our approach thus differs from Ip et al. (2020) and Petters and Schroeder (2019) in several respects: (i) the fact that we explicitly elicit the fairness perception of affirmative action policies, (ii) the nature of the affirmative action policies under study, and (iii) the context and outcomes we consider (willingness to compete, output produced, post-competition cooperation in teamwork and retaliation).

We introduce an experimental design allowing us to quantify productivity, working time, and luck separately and precisely within a unified framework. Each subject participates in two sessions, conducted in two consecutive weeks. We measure productivity and choice of working time in the preparatory session of the experiment. In a real effort task, we first measure subjects' productivity and their individual choice of working time. Luck is a randomly assigned multiplier, which upgrades output of lucky subjects, but downgrades output of unlucky subjects, discriminating against the latter group.

In the main session, we build on the design of Balafoutas and Sutter (2012), yet differ significantly in the criteria for affirmative action and add several outcome variables. In our design, affirmative action policies are based on the three determinants of performance. We vary the rule determining winners of the tournament between treatments. In the control treatment, the winners are the two subjects with the highest performance. In the luck/working time/productivity treatments, at least one of the two winners must be a subject that is unlucky, has a short working time, or low productivity, respectively. We argue that differences in the consequences of affirmative action across treatments are due the criteria (luck, or productivity, or working time) this policy is based on.

Our study is the first to analyze the consequences of affirmative action policies related to productivity and working time. We are also able to systematically compare affirmative action policies based on the three determinants of performance in a unified framework. Moreover, we provide novel evidence on heterogeneity in the perceived fairness of affirmative action policies and link fairness perceptions to its consequences. Our finding that the fairness perception of an affirmative action policy can shape its consequences is key for the communication and successful implementation of such policies if they are politically desired.

The remainder of the paper is organized as follows. Section 2 explains our experimental design, while Section 3 presents our results on the consequences of affirmative action on willingness to compete, efficiency, cooperation, and retaliation and provides evidence that subjects internalize affirmation action policies beyond the context in which they are applied. Section 4 discusses our findings and concludes.

2 Experimental design

Our experiment employs a combination of a within- and between-subject design, in which the four treatments are assigned across subjects: one control treatment without affirmative action and three treatments with different affirmative action policies. Each affirmative action policy favors subjects with a characteristic that dampens their performance – either subjects who have bad luck (which resembles discrimination), those with low productivity, or those who have chosen a short working time.

Each subject participates in two sessions taking place in consecutive weeks: a preparatory session and the main session. The purpose of the preparatory session is to learn about each subject's productivity and individually chosen working time to classify them into high and low productivity, high and low working time types, respectively. This determines which subjects will be favored by the respective affirmative action policies. In the main session, we investigate the fairness perception of the various affirmative action policies and their consequences on willingness to compete, efficiency, and post-competition cooperation and retaliation. Table 1 provides an overview of the experimental design.

Preparatory session		Practice round (grid task)
		Measurement of baseline productivity (grid task)
		Questionnaire
		Measurement of choice of working time (grid task)
Main session	Stage 1	Piece rate (grid task)
	Stage 2	Tournament (grid task)
	Stage 3	Choice between piece rate and tournament (grid task)
	Stage 4	Group work (slider task)
	Stage 5	Dictator Game
		Measurement of fairness perception

Table 1. Summary of experimental design.

2.1 The real-effort task

We apply the different affirmative action policies to performance in a tournament that is based on what we call the *grid task*, a real-effort task introduced by Abeler et al. (2011). Subjects work on this task several times under different incentive schemes (see Table 1). In this task, subjects count the number of zeros in a 10-by-10 table containing 100 digits of randomly distributed zeros and ones (see Figure 1).

The grid task has several desirable attributes. First, the tediousness of the task induces a positive effort cost and minimizes experimenter demand effects (Abeler et al., 2011). Therefore, we are confident that our measure of working time (the time subjects decide to work on this task) actually captures the effort subjects are willing to spend on the task. Second, our data show substantial variation in productivity and chosen working time for this task (see Figure A.1 and Figure A.2 in Appendix A). Third, the grid task does not require special prior knowledge or skills. Moreover, as Balafoutas et al. (2016) note and our data confirm, the grid task is a gender-neutral one. For example, the average number of correctly solved grids in the five minute grid task in the preparatory session is 7.46 for men and 7.69 for women (Mann-Whitney U test, $p = 0.232$).[5] Most importantly, the task allows us to clearly distinguish between the three determinants of performance that matter for perceived distributional fairness according to different fairness ideals (see, e.g., Cappelen et al., 2007).

[5] Considering only those subjects who participated in both sessions, as we do in Section 3, these numbers hardly change (7.54 and 7.68 correctly solved grids for men and women, respectively; not significantly different according to Mann-Whitney U test: $p = 0.478$).

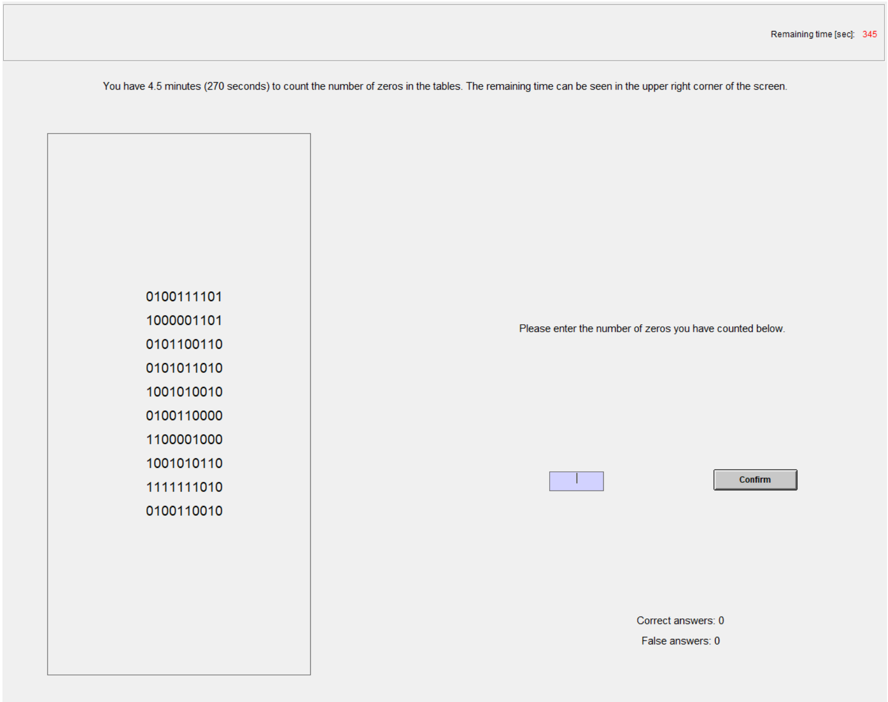

Figure 1. Exemplary screen of grid task

2.2 Preparatory session

The preparatory session consists of four parts (see Table 1). First, subjects familiarize themselves with the grid task in a practice round of two minutes. We then measure each subject's productivity, followed by a questionnaire and a choice of individual working time. The purpose of this is to classify subjects into binary types with high and low productivity or long and short working time, respectively. Depending on the treatment allocation, a subject's type determines whether a subject will be among those favored by the affirmative action policy in the main session.

Classification into productivity type: Subjects are asked to solve as many grids as they can within five minutes at a piece-rate of 0.50 EUR for each correctly solved grid. Subjects who solve more grids correctly than the median are classified as being of the high productivity type, while those below the median are classified being of the low productivity type.[6] Figure A.1 in Appendix A in the appendix displays the distribution of the number of correctly solved grids in this stage.

Classification into working time type: At the end of the preparatory session, subjects work on the grid task for another time, now at a piece-rate of 0.10 EUR per correctly solved grid. Subjects can now freely choose how long they want to work. After

[6] Our measurement of productivity (number of correctly solved grids per minute) might capture both the given ability of working on the grid task and effort at the intensive margin. However, subjects tend to exert maximum effort in laboratory real effort tasks with short working period (Araujo et al., 2016; Corgnet et al., 2015; Gächter et al., 2016; Goerg et al., 2019). Therefore, our preferred interpretation of productivity is that it reflects ability at the grid task, which is beyond subjects' control during the experiment.

every grid, subjects can choose to continue or stop working by clicking on the corresponding button. If subjects choose to stop working, they finish the preparatory session and can leave the laboratory immediately. To minimize peer effects in the decision when to stop working, we implement a flexible show-up policy, meaning that subjects start the session individually and do not reach this last stage at the same time.[7] We truthfully communicate to subjects upfront that their chosen working time has additional consequences on the session in the following week. Based on whether their chosen working time is below or above the median, they are classified as being either of the short or long working time type. This determines how long they will work on the grid task in the main session and has consequences on their expected earnings in the main session.[8] The low piece-rate was deliberately chosen to make the task less attractive so subjects would choose to stop working after a reasonable time. On average they do so after 24.13 minutes.[9] Figure A.2 in Appendix A displays the distribution of working time. In our data, the number of correctly solved grids in the first stage and chosen working time in the last stage are not significantly correlated (Pearson's correlation coefficient: $\rho = 0.061$, $p = 0.192$).[10]

Questionnaire: We elicit a number of control variables, including measures of risk and social preferences, cognitive ability (Raven matrices), personality (Big Five), and socio-demographics. Appendix D provides more details on the questionnaire.

2.3 Main session

The main session consists of five stages and a final questionnaire.[11] In Stages 1 to 3, subjects work on the grid task repeatedly, with their payoff-relevant performance being determined as follows:

[7] In the recruitment e-mail, subjects are informed that they can show up at the lab at any time within a two-hour interval.

[8] Refer to Section C.1 in the appendix for details on how the consequences on the working time decision were communicated. In particular, the instructions state "Your working time today will determine your working time in the next session next week. Next week, you will work on a similar task and you will be given a specific amount of time to solve as many tables as possible and get paid accordingly. (...) Based on your chosen working time today, we will form two groups. One group contains that half of the participants who choose to work for a shorter time today. This group will also be given a shorter time to work in the session next week. The other group contains that half of the participants who choose to work for a longer time today. This group will also be given a longer time to work in the session next week. Who works shorter will, on average, solve less tables correctly and therefore earn less. The experiment is, however, shorter (it will end earlier). Who works longer will, on average, solve more tables correctly and therefore earn more. The experiment is, however, longer (it will end later)."

[9] Considering only those who participated in both sessions the average time after which subjects stop working is 24.17 minutes.

[10] Considering only those who participated in both sessions we find a correlation coefficient of $\rho = 0.034$ ($p = 0.513$).

[11] Part of the design of the main session builds on Balafoutas et al. (2016). We thank the authors for sharing their ztree program and instructions with us.

Performance = Correct grids per minute × Working time × Luck multiplier

Assignment of types. In our experiment, each subject is of one of eight (2^3) types: high or low productivity × long or short working time × lucky or unlucky. Subjects are fully informed about all three dimensions of their own type before they enter the first stage of the main session and a subject's type stays constant throughout the experiment. The *productivity* and *working time type* are assigned to each subject based on the outcomes of the preparatory session as described in Section 2.2 above. While productivity is something given, subjects are given more or less time to work on the grid task in the main session according to their chosen working time in the preparatory session. Subjects of the long working time type are given 7.5 mins to work on the task, while those of the short working time type have only 4.5 mins. *Luck* is reflected by a randomly assigned multiplier. Half of the subjects are *lucky*. They are assigned a high multiplier of 1.25. The other half are *unlucky* being assigned a low multiplier of 0.75. The number of correctly solved grids in the total time worked is weighted with multiplier.[12] Those parameters were chosen to make the effects of each policy on the probability of winning for those favored (not favored) by it comparable in size.[13]

Stage 1: Piece Rate. This stage provides a baseline measure of performance without tournament incentives. Subjects work on the grid task according to their type's working time and receive a piece-rate payment of 0.50 EUR for each correctly solved grid multiplied with the respective luck multiplier.

Stage 2: Tournament. In stage 2, subjects solve the grid task under tournament incentives, each competing against five other subjects.[14] The purpose of this stage is to measure the effect of different affirmative action schemes on performance. Among each group of six, the two winners of the tournament receive 1.50 EUR per correctly solved grid multiplied with their luck multiplier, while losers receive noth-

[12] For example, if a subject has solved 12 grids correctly and has been assigned the high (low) multiplier, the 12 grids are treated as 15 (9) grids. By randomly assigning the luck multipliers at the beginning of the main session, and having them remain constant for each subject throughout the experiment, we effectively discriminate against unlucky individuals.

[13] Specifically, after having observed the productivity distribution in the preparatory session, we run a simple simulation that determines how likely it is on average for the low and high productivity type to win the tournament with and without affirmative action regarding productivity. We then choose parameters for the working time of the short and long working time type such that the change in the average probability of winning the tournament when introducing affirmative action regarding working time is comparable. This implies that the luck multiplier for unlucky and lucky subjects is chosen such that the change in the average probability of winning the tournament when introducing affirmative action regarding luck is also comparable.

[14] A potential confound of the tournament outcome could arise from an unbalanced composition of types across groups because the probability of winning the tournament depends both on one's own and others' performance, which in turn is affected by the types of competitors. To eliminate this effect, each group of six consists of three subjects with low productivity, three with high productivity, three with long working time, three with short working time, three with the high and three with the low luck multiplier. At the beginning of stage 2, subjects are informed about this rule for group composition, but not about the specific type of other each group member. Group composition remains the same in all following stages.

ing, keeping the average payment constant compared to stage 1. If necessary, a random tie-breaking rule is applied to determine winners. Winners and rank within each group of six are not announced until the end of stage 5. How exactly the winners are determined is the main treatment variation of our experiment.

Between-subject treatments. We conduct four treatments: one control treatment without affirmative action and three different affirmative action (AA) treatments. In the control treatment without any quota rule, the two subjects with the highest performance are the winners. In the affirmative action treatments, a quota rule is added to determine the winners. If this rule is not automatically fulfilled, the subject with the second-highest performance is replaced by the highest-performing subject who fulfills the quota criterion. The following quota rules apply in the treatments:

- *Control (CTR)*: No quota rule.
- *Affirmative action w.r.t. luck (AAL)*: At least one subject of the unlucky type has to be among the two winners.
- *Affirmative action w.r.t. working time (AAW)*: At least one subject of the short working time type has to be among the two winners.
- *Affirmative action w.r.t. productivity (AAP)*: At least one subject of the low productivity type has to be among the two winners.

Control questions make sure that subjects understand the tournament scheme before starting to work on the task.[15]

Stage 3: Self-selection into tournament. To elicit the willingness to enter the tournament, subjects work on the grid task again. In this stage, they choose whether they would like to work under piece-rate incentives (exactly as in stage 1) or tournament incentives (as in stage 2). Importantly, if a subject chooses the tournament in *stage 3*, her performance will be compared to the performance of her five fellow group members in *stage 2*. This feature ensures that a subject's decision to enter the tournament is independent of her belief about others entering the tournament (compare Niederle et al., 2013).

Belief elicitation. At the end of stage 3, subjects report their beliefs about their relative performance in stage 1, 2, and 3. Subjects are asked to guess their rank both within the whole group of six and within the group of three subjects with the same luck type (in treatment AAL), the same working time type (in treatment AAW), or the same productivity type (in treatment AAP). One guess is randomly chosen to be payoff-relevant. Subjects receive 1 EUR if they guess correctly.

[15] After having read the rules of stages 2 and 3, subjects have to answer control questions correctly before they can start working on the grid task. These multiple-choice questions describe scenarios about competition within a group, provide information about each member's performance and who is favored by affirmative action, and ask about the winners. The control questions cover both cases in which affirmative action does or does not change the results of the competition.

Stage 4: Cooperation in group work. Stage 4 keeps the group composition and treatment history from previous stages, but provides a new working environment with a new task and new payoff rules. Compared to previous stages, all subjects now work for the same amount of time (5 mins), and there are neither multipliers nor affirmative action.

In the slider task (Gill and Prowse, 2012), subjects are shown a series of screens, each with 6 sliders on them. Each slider has a range of positions between 0 and 100. Sliders are solved by using the computer mouse to move the slider markers to the position of 50.[16] A screen is considered "solved" if all six sliders are positioned at 50. Only then can a subject continue to the next screen.

Importantly, each correctly solved screen yields 0.60 EUR *for the group as a whole*, 0.10 EUR for each of its members. Since all group members benefit from an individual's effort, this is a typical setup to measure cooperation and how much a subject works indicates their willingness to contribute for the benefit of the group. In order to keep the previous tournament experience with or without affirmative action salient, we introduce unequal bonuses for winners and losers of the tournament in stage 2 (as Balafoutas et al., 2016, do). Subjects receive a bonus of 5 EUR if they were the winners in stage 2, and 2 EUR otherwise.

Stage 5: Dictator Game. Subjects play one Dictator Game with each of their five group members. The only thing they know about the other group members is whether they were winners in stage 2 and whether they were favored by affirmative action. All five Dictator Games are displayed on the same screen. For each game, subjects are endowed with 5 EUR, and can decide how much to give away in 0.1 EUR increments. This setup is used to learn how more or less favorably subjects treat specific other subjects after the tournament phase.

Fairness perception. After stage 5, subjects are asked how fair they perceive the different policies to be. They first rate the policy in their own treatment on a seven-point Likert scale, then the policies that appear in the other treatments. Thus, we assess the fairness perception of each subject for each of the four policies. The fairness questions describe the policies neutrally and do not mention the term "affirmative action" (see Appendix B for the exact wording).

2.4 Procedures

The experiment was conducted at the DICE Lab at the University of Düsseldorf in April 2018 and the BonnEconLab at the University of Bonn in August 2018 using the software zTree (Fischbacher, 2007). On average, each session lasted 90 minutes. Subjects were recruited via ORSEE (Greiner, 2004) and Hroot (Bock et al., 2014) from the subject pools of the respective labs, both of which include students of various disciplines. In the recruitment email, subjects are informed that the experiment consists of two mandatory sessions and that all payments will only be realized at

[16] To make sure that subjects use only the computer mouse to solve the task, the left and right arrow keys of the keyboard are disabled.

the end of the second session. Only 7 out of 463 subjects who participated in the first preparatory session did not show up in the second one (main session), implying an attrition rate of 1.5%. To be able to match the data of both sessions while ensuring anonymity, we asked subjects to generate an ID (that is never connected to their name) at the beginning of the preparatory session, and to re-enter it in the main session. In total, the number of subjects in each treatment is 108, 84, 90, and 90 for CTR, AAL, AAW, and AAP, respectively.[17] Throughout the paper, we focus our analysis on those subjects who participated in both sessions unless explicitly stated otherwise. Gender composition does not vary significantly between treatments (59% females; Kruskal-Wallis test: $p = 0.582$).

On average, subjects earned 26 EUR for both sessions. The payoff of the preparatory session is the sum of payoffs from the measurements of productivity and choice of working time in the grid task, a risk choice list, and a fixed payment of 2 EUR for completing the questionnaire. The payoff of the main session consists of a 4 EUR show-up fee and the earnings from one randomly chosen stage. Subjects know that one of the five parts of the main session will be randomly chosen to be paid. Instructions are distributed stage-by-stage. At the end of each of the first four stages, subjects are informed about their individual performance. Subjects never learn the performance of other subjects.

3 Results

We start by presenting evidence on the fairness perception of the various affirmative action policies. We then analyze the consequences of affirmative action for willingness to compete in the tournament and efficiency, before we examine its impact on post-competition cooperation and retaliation. Finally, we provide evidence that affirmative action policies are internalized beyond the context in which they are binding. Throughout, we discuss links between fairness perception and consequences of affirmative action policies.

3.1 Fairness perceptions of affirmative action

Based on fairness considerations, there is broad support for all affirmative action policies under study (see Figure 2). None of them is considered less fair than an absence of affirmative action. Affirmative action in favor of unlucky subjects (AAL) is perceived as fairest, followed by affirmative action in favor of subjects of the low working time type (AAW), while affirmative action favoring subjects of the low productivity type (AAP) is rated roughly equal to no affirmative action (CTR). The average scores on a Likert scale from 1 to 7 are 4.70 for AAL, 4.23 for AAW, 3.31 for AAP, and 3.25 for CTR.[18] Wilcoxon signed rank tests for all pairwise policy compar-

[17] To guarantee a similar composition of all groups (see description of Stage 2: Tournament), we have to exclude around 18% of subjects who showed up from participating in the main session.

[18] Figure A.3 in the appendix displays the distributions of answers.

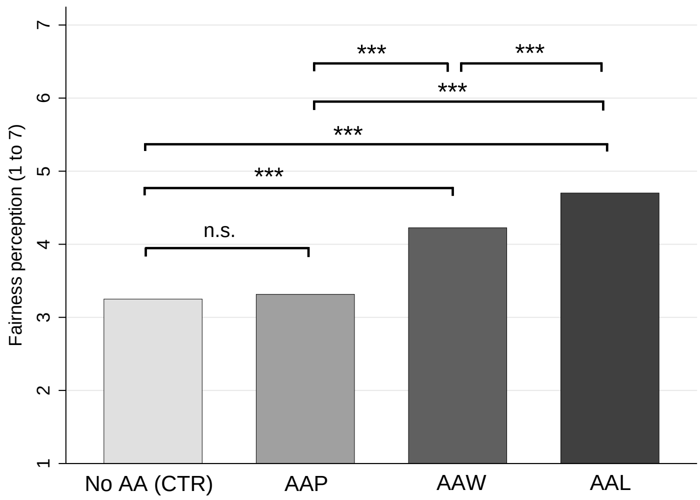

Figure 2. Fairness perception

Notes: Fairness perceptions for different policies. Higher numbers indicate that a policy is perceived as fairer. The brackets and stars above each bar show results of Wilcoxon signed rank tests. *** indicates $p < 0.001$.

isons yield $p < 0.001$.[19] The only exception is the absence of a significant difference in perceived fairness of treatments CTR and AAP (Wilcoxon signed rank test, $p = 0.327$). Interestingly, the ranking of treatments by perceived fairness diverges slightly from meritocracism, according to which a policy controlling for an out-of-personal-control factor (productivity) is rated as fairer than an in-control factor self-chosen working time.

Result 1 (Fairness perception): On average, affirmative action policies that compensate for bad luck (discrimination) or short working time are perceived as significantly fairer than no affirmative action. Affirmative action based on low productivity is perceived as equally fair as no affirmative action.

This pattern of fairness perceptions is similar across treatments (see Figure A.4 in the appendix). Moreover, the average fairness rating of favored and unfavored subjects does not differ significantly.[20] Importantly, these findings jointly underline that fairness judgments regarding affirmative action are not strongly shaped by individual experiences.

In sum, different affirmative action policies are perceived as differently fair. Independent of which policy is actually in use, AAL and AAW are perceived as fairer than AAP or no affirmative action (CTR).

[19] Throughout the paper, we report two-sided tests unless explicitly stated otherwise.

[20] An OLS regression of all 1056 fairness ratings from 264 subjects in the three treatments with affirmative action on a binary variable indicating whether a subject is favored yields a coefficient of -0.072 with a p-value of 0.603 (standard errors clustered at subject level).

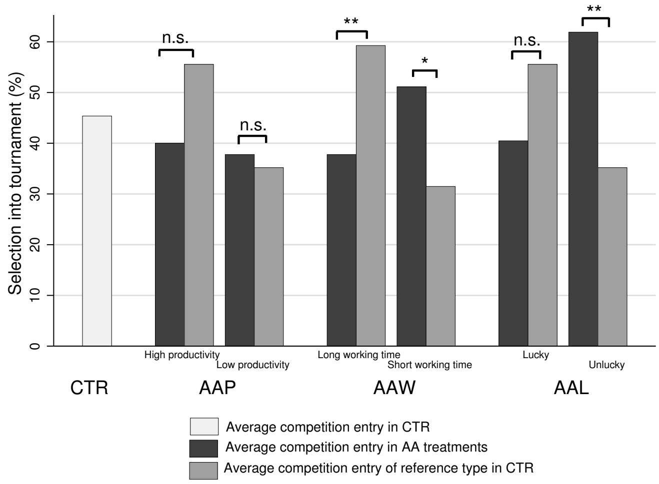

Figure 3. Willingness to compete by treatment and type

Notes: Proportion of competition entry in stage 3 of a given type in the treatments with and without affirmative action concerning their type. The brackets and stars above each bar show results of Fisher's exact tests, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

3.2 Willingness to compete

A key purpose of affirmative action policies is to encourage those favored by them to enter the competition. In this section, we will analyze whether the different policies indeed affect subjects' willingness to select into the tournament when given the choice between working under a piece rate and a tournament scheme.

In Section 3.2.1 we start by comparing the average willingness to compete in stage 3 across treatments and then take a closer look at favored and non-favored types. Throughout, we will point out close links between fairness perceptions and willingness to compete. In Section 3.2.2 we consider determinants of willingness to compete at the individual level.

3.2.1 The big picture: results at the aggregate level. Overall, the proportion of subjects selecting the tournament differs modestly across treatments. 45% of subjects are willing to compete in the absence of affirmative action (CTR) compared to 39% in AAP, 44%, in AAW, and 51% in AAL. For the affirmative action schemes, the overall pattern mimics that in Figure 2: subjects tend to be more willing to compete under affirmative action schemes that are generally perceived as fairer. However, differences in willingness to compete are not significant across treatments.[21]

Since ultimately, we want to learn about the effects of affirmative action on those who are favored by it compared to those who are not, this result warrants closer inspection. We thus continue by comparing the willingness to compete of

[21] Fisher's exact tests for pairwise comparisons with the control treatment yield $p = 0.388$ for AAP, $p = 1.000$ for AAW, and $p = 0.468$ for AAL. Likewise, a Kruskal-Wallis test does not detect any significant differences across all four treatments ($p = 0.446$).

each type (e.g., the unlucky type) in the treatment that concerns them (e.g., AAL) to that of the same type in the control treatment (see Figure 3).

We observe a similar tendency in AAL and AAW. While affirmative action increases favored subjects' willingness to compete, it tends to lower tournament entry of non-favored subjects. The encouragement effect is most striking for AAL, in which unlucky subjects' tournament participation increases by 26.7 percentage points compared to CTR (Fisher's exact test, $p = 0.013$). Subjects of the short working time type are 19.6 percentage points more likely to compete in AAW than in CTR (Fisher's exact test, $p = 0.064$). However, in both treatments, we also observe a discouragement effect of 15.1 and 21.5 percentage points, respectively, on non-favored subjects (not significant in the case of AAL; Fisher's exact tests, $p = 0.156$ for AAL and $p = 0.044$ for AAW).

In contrast, in AAP tournament entry of low productivity subjects hardly increases compared to CTR (2.6 percentage points; Fisher's exact test, $p = 0.836$), while high ability subjects tend to be discouraged from entering (15.6 percentage points, Fisher's exact tests, $p = 0.158$).

Interestingly, the encouragement effect of affirmative action on the favored subjects closely mirrors the average fairness rating of the affirmative action treatments (compare Figure 2) with higher encouragement in treatments that are generally perceived as fairer.

Result 2a (Willingness to compete): The higher the average fairness rating of an affirmative action scheme, the more willing are favored subjects to compete.

Result 2a establishes a relationship between the average fairness perception and the effectiveness of affirmative action to increase its target group's willingness to compete. A higher average fairness perception of an affirmative action scheme goes hand in hand with a higher willingness to compete of the favored subjects. A plausible explanation for such a relationship is that subjects anticipate the social acceptance, i.e., the average fairness perception of a given quota rule, and feel more comfortable to enter a competition, in which they are favored if it is generally judged as fair. Of course, average perceived fairness of an affirmative action scheme also reflects personal perceived fairness of that scheme. In the following, we will investigate the relationship between fairness perception and willingness to compete at the individual level.

3.2.2 Determinants of the willingness to compete at the individual level. Table 2 reports the marginal effects resulting from probit regressions on the determinants of willingness to compete at the individual level. The dependent variable is an indicator variable that equals 1 if a subject chooses to enter the tournament and 0 if they choose the piece rate payment scheme in stage 3. From a theoretical perspective, an individual's belief on the likelihood of winning the tournament and risk attitude should be the key determinants of this choice. Empirically, there is

	(1)	(2)	(3)
Belief of rank	-0.100***	-0.088***	-0.119***
	(0.015)	(0.017)	(0.017)
Risk attitude	0.026**	0.029**	0.031***
	(0.010)	(0.012)	(0.012)
Female	-0.093*	-0.104*	-0.091*
	(0.049)	(0.057)	(0.055)
Fairness perception	0.019	0.028*	0.049**
	(0.013)	(0.015)	(0.021)
Favored			0.480***
			(0.129)
Fairness perception × Favored			-0.067**
			(0.029)
N	372	264	264

Notes: Average marginal effects from a probit regression. Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The binary dependent variable is willingness to compete in stage 3 (1 if tournament, 0 if piece-rate). *Belief on rank* is a categorical variable about beliefs regarding own rank (between 1 and 6) in stage 3; *Risk attitude* is the answer to the general risk question elicited on an 11-point scale, higher numbers indicate a higher willingness to take risks; *Female* is an indicator variable for gender (1 if female, 0 if male); *Fairness perception* reflects fairness rating of own treatment, elicited on a 7-point scale on which higher numbers indicate higher perceived fairness; *Favored* is an indicator variable (1 if favored, 0 otherwise).

Table 2. Individual level determinants of willingness to compete

also substantial evidence that women are less likely to compete (Almås et al., 2016; Buser et al., 2014; Croson and Gneezy, 2009; Gneezy et al., 2003; Leonard et al., 2009; Niederle and Vesterlund, 2007, 2011; Sutter and Glätzle-Rützler, 2015). Given the results in Section 3.2.1 we are interested in the role of the fairness perception of the policy in place.

Results in column (1) imply that the belief regarding their own rank in the tournament, risk attitudes, and gender are indeed predictive for individual willingness to compete. If a subject's belief about their own rank increases by 1 (on a scale from 1 to 6, with 1 as the highest rank), it is about 10 percentage points less likely to enter the tournament.[22] If their willingness to take risks increases by 1 (on an 11 point Likert scale), they are 3 percentage points more likely to compete. Women tend to be 9 percentage points less likely to compete compared to men. Finally, on average an individual's own fairness perception is not a significant predictor of willingness to compete. We move on by investigating whether the impact of fairness perceptions differs for favored and non-favored subjects (see column (3)). Since this requires

[22] Table A.1 replicates all specifications of Table 2 and shows the marginal effect of each value of belief separately. Moreover, results are qualitatively the same and quantitatively very similar if we use beliefs regarding own rank in stage 2 instead of stage 3.

dropping subjects from the control treatment, column (2) serves as a reference, replicating column (1) with subjects in the affirmative action treatments.

Results in column (3) imply that an individual's own fairness perception is a highly significant and influential predictor of the non-favored subjects' willingness to compete: if their fairness perception of a given affirmative action treatment increases by 1 (on a 7 point Likert scale), they are about 5 percentage points more likely to compete. Moreover, the specification reported in column (3) reveals that favored subjects tend to have a higher willingness to compete but their individual fairness perception of an affirmative action scheme does not have any effect on their willingness to compete. A test of joint significance of Fairness perception and Fairness perception × Favored yields $p = 0.429$. Finally, women are slightly less likely to compete than men.

Result 2b (Willingness to compete): At the individual level, willingness to compete is driven by an individual's belief of winning and risk attitude. Moreover, non-favored subjects are more likely to compete, the fairer they consider an affirmative action scheme to be, while this is not the case for favored subjects.

Overall, a higher fairness perception of affirmative action schemes increases the share of competing subjects in two ways: a higher *general* fairness perception of a given quota rule goes hand in hand with a higher willingness to compete of favored subjects and a higher *individual* fairness perception raises non-favored subjects' willingness to compete.

While encouraging favored individuals to enter competitions is the key aim of affirmative action, raising non-favored subjects' willingness to compete is also desirable since competition tends to increase efficiency (see, e.g., Balafoutas et al., 2016, Niederle and Vesterlund, 2007, and Figure 4). Thus, our results point at an important aspect in designing and communicating affirmative action schemes that has so far been disregarded, namely their fairness perception. For an affirmative action policy to encourage favored subjects to compete without discouraging non-favored subjects, it is vital that the policy is perceived as fair in general (by others) and personally. So from a policy perspective, providing a convincing rationale for the implementation of quota rules to ensure that they are largely perceived as fair seems key to make them a success. An example would be providing evidence on the discrimination against women when introducing a gender quota (Ip et al., 2020).

3.3 Efficiency

A prominent worry of opponents of affirmative action is that it harms efficiency, i.e., that it results in not selecting or rewarding the "best". A 50% quota rule that replaces the second-highest performer as a winner by someone else may seem damaging to efficiency from an ex-ante perspective. But at closer inspection this is not necessarily true.

To learn about efficiency we use the tournament winners' number of correctly solved grids per minute as a measure, thus not considering differences in performance that arise due to luck or working time. This measure actually captures two separate aspects. The first is who the winners are and would be of most interest in a promotion or selection context. This is determined both by winner selection according to the quota rule (or its absence) and participants' entry decision. The second determinant is how well winners perform under the given incentives. This is likely driven by motivation and more important when output during the tournament matters.[23]

Qualitatively, results are the same, no matter whether we consider the number of correctly solved grids per minute of stage 3 winners in stage 3 or in the preparatory session as an efficiency measure (see Figure 4): the number of correctly solved grids per minute does not differ significantly between the control treatment and those with affirmative action (test results reported in the figure notes).

Result 3 (Efficiency): None of the affirmative action policies under study harms efficiency.[24]

This result extends the findings of Balafoutas and Sutter (2012) and Niederle et al. (2013) that a 50% quota rule favoring women does not harm efficiency to 50% quota rules that favors individuals with bad luck, short working time or low productivity, i.e. it holds regardless of which determinant of lower performance affirmative action compensates for.

Finally, how fair an individual perceives the assigned treatment to be does not correlate with the number of correctly solved grids per minute in stage 3, neither for winners of the tournament nor for subjects as a whole ($\rho = -0.073$, $p = 0.554$ for winners, $\rho = -0.037$, $p = 0.482$ for all subjects).

3.4 Post-competition cooperation, retaliation, and internalization of social norms

Cooperation. We measure cooperative behavior by performance in the slider task (number of correctly solved slider screens) in stage 4 in which a higher individual performance yields equal benefits for all members of a group. Overall, average post-competition cooperation at the group level does not differ significantly across treatments (Kruskal-Wallis test, $p = 0.308$). Neither do we find significant differences for pairwise comparisons between cooperation levels in each affirmative action treatment and the control treatment (see Table A.2 in the appendix; Mann-Whitney U tests $p > 0.10$ for all). Importantly, the lack of significant differences is not due to a lack of statistical power. Our sample size is able to detect an effect size as small as 0.87 (roughly a 12% performance change based on the con-

[23] For a more detailed discussion of these separate aspects refer to Section A.4 in the Appendix.
[24] Extending the efficiency analysis to the performance of both winners and non-winners, we also observe that none of the affirmative action policies harms efficiency.

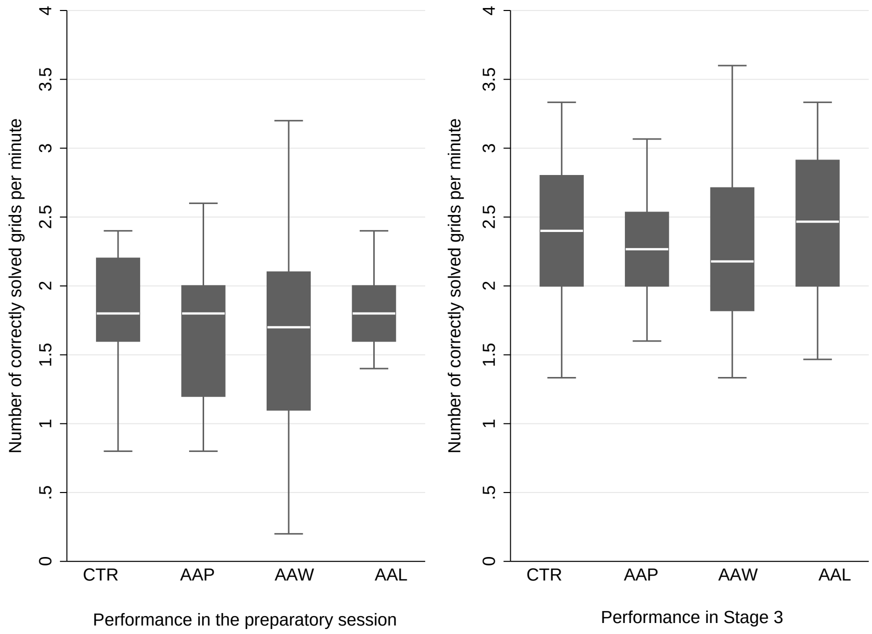

Figure 4. Boxplots of the number of correctly solved grids per minute in the preparatory session and stage 3 of winners in stage 3

Notes: Number of correctly solved grids per minute of stage 3 winners in the productivity task of the preparatory session and stage 3 for each treatment. Two outliers from CTR and AAP in the right boxplot are excluded. The upper (lower) hinges of the boxes show the 75th (25th) percentiles, the white lines inside the boxes show the median values, and the upper (lower) adjacent lines show the maximum (minimum). Stage 3 winners solved on average 1.83 grids per minute correctly In the preparatory session in the control treatment compared to 1.85, 1.73 and 1.76 in treatments AAP, AAW, and AAL, respectively. The corresponding p-values of Mann-Whitney U tests comparing the affirmative action treatments to the control treatment are 0.815, 0.554, and 0.953, respectively. In stage 3, the winners solved 2.47 grids per minute correctly in the control treatment compared to 2.51, 2.32, and 2.41 in treatments AAP, AAW, and AAL, respectively (p-values of Mann-Whitney U tests are 0.642, 0.390, and 0.920, respectively).

trol treatment) at the conventional level of power of 80% in each treatment using a t-test and a significance level of 0.05. This confirms the results of Balafoutas et al. (2016), Sutter et al. (2016) and Kölle (2017) who find no spillover effects of quotas on subsequent teamwork in affirmative action contexts that differ from the context we study.

Furthermore, we find no evidence that fairness perceptions affect post-competition cooperation (Pearson's correlation coefficient between fairness perception of own treatment and performance in the slider task: $\rho = 0.047$, $p = 0.382$).

Retaliation. In contrast to the slider task, in which an individual's behavior affects all other group members equally, decisions in stage 5's dictator games allow subjects to treat each individual other group member in a more or less favorable way. They may condition their transfer on whether someone was a winner or loser

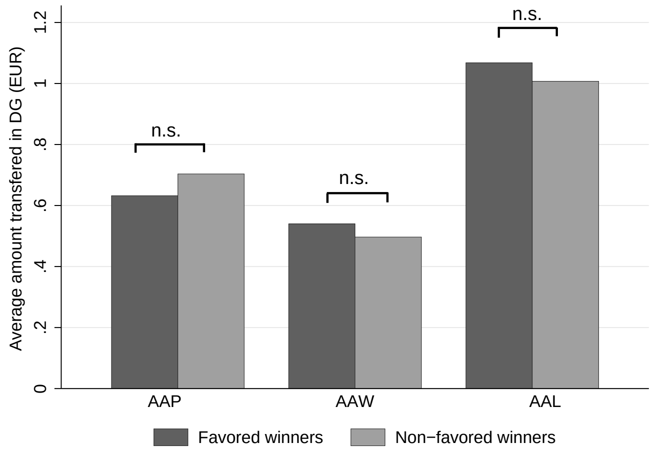

Figure 5. Retaliation in dictator games

Notes: Average transfer (EUR) from non-favored losers to favored winners versus non-favored winners in dictator games for each treatment. The brackets and stars above each bar show results of Wilcoxon signed rank tests, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

and favored or non-favored by the respective affirmative action policy in the tournament in stage 2.[25]

In particular, non-favored subjects may have an urge to treat formerly favored subjects in a less advantageous manner than non-favored ones ("retaliation"). Our analysis focuses on the most pointed situation, in which such retaliation seems most likely to occur: we investigate whether non-favored losers give less to favored winners than to non-favored winners in the affirmative action treatments (see Figure 5). We find no evidence for such retaliation under any of the three affirmative action policies (Wilcoxon signed rank tests yield $p = 0.198$ for AAL, $p = 0.906$ for AAW , and $p = 0.317$ for AAP, respectively). Figure A.5 in the appendix provides additional evidence on the absence of retaliation based on a broader set of situations.

While there is no evidence for retaliation at the aggregate level, it is worth noting that non-favored subjects' transfer less to winners if they perceive an affirmative action policy as less fair (see marginal effects of a tobit regression displayed in Table 3: Panel 1). This finding holds both for the group of non-favored subjects overall as well as for the subgroup of non-favored losers. Thus, the absence of evidence on retaliation at the aggregate level masks the fact that winners are retaliated against when affirmative action is perceived as less fair, while they receive higher transfers if it is perceived as fair. However, subjects do not distinguish between favored and non-favored winners.

By pointing out this important role of a policy's fairness perception for its impact on post-competition interactions, our results document a link between previous findings on backlash against favored individuals (e.g., Fallucchi and Quercia,

[25] Considering the overall average transfer in the dictator games as an indicator, we do not observe any significant differences in pro-social behavior across treatments.

from	Panel 1: Non-favored losers				Panel 2: All subjects			
to	fw	nfw	fl	nfl	fw	nfw	fl	nfl
Fairness	0.262**	0.211*	0.101	0.089	0.203**	0.099	0.044	0.018
	(0.122)	(0.120)	(0.102)	(0.090)	(0.083)	(0.079)	(0.063)	(0.058)
Constant	-0.966*	-0.736	0.444	0.551	-0.654*	-0.336	0.741***	0.718***
	(0.520)	(0.508)	(0.403)	(0.352)	(0.367)	(0.352)	(0.274)	(0.250)
N	89	86	89	89	221	215	263	264

Notes: Tobit regressions, standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable is the average transfer (0 to 5 EUR) in dictator games (stage 5). *Fairness* reflects fairness perception of the treatment policy, ranging from 1 (completely unfair) to 7 (completely fair). "fw" - "favored winners", "nfw" - "non-favored winners", "fl" - "favored losers", "nfl" - "non-favored losers".

Table 3. Transfer in dictator games and fairness perception

2018; Leibbrandt et al., 2017) and support for affirmative action policies (Ip et al., 2020).

Result 4 (Post-competition cooperation and retaliation): Overall, the affirmative action policies under study do not affect post-competition cooperation and retaliation. However, winners are retaliated against when affirmative action is perceived as less fair and receive higher transfers if it is perceived as fair.

Internalization of affirmative action norms. Finally, our data on post-competition redistributive behavior in stage 5 also allow us to investigate whether subjects "internalize" the normative content of affirmative action policies beyond the context in which they are applied. Intuitively, affirmative action policies aim at improving outcomes for favored individuals. Thus, they convey the social norm that specific individuals should be treated preferentially. In the case of favored losers, this aim has not been achieved. Even though they have been encouraged to compete, they still received the bad outcome (i.e., lost the competition).

In such a situation, subjects who have internalized the norm conveyed by the affirmative action policy could decide to use other available means to support the policy's target group. In the context of the stage 5 dictator games, internalization would imply transferring more to favored than non-favored losers.

As displayed in Figure 6, we indeed find evidence in favor of internalization of affirmative action norms that is closely linked to average fairness perception. In treatments AAL and AAW that are rated as significantly fairer than AAP and the control treatment, subjects transfer 23% and 15% more to favored than non-favored losers in the stage 5 dictator games (Wilcoxon signed rank tests: $p < 0.001$ and $p = 0.001$, respectively). This is not the case in treatment AAP (Wilcoxon signed rank test, $p = 0.412$) that is not perceived as fairer than an absence of affirmative action. While we observe that subjects treat favored losers preferentially in exactly those treatments that are generally perceived to be fairer, we do not find a significant relationship with individual perceived fairness of the given treatment (see Table 3: Panel 2, right two columns)).

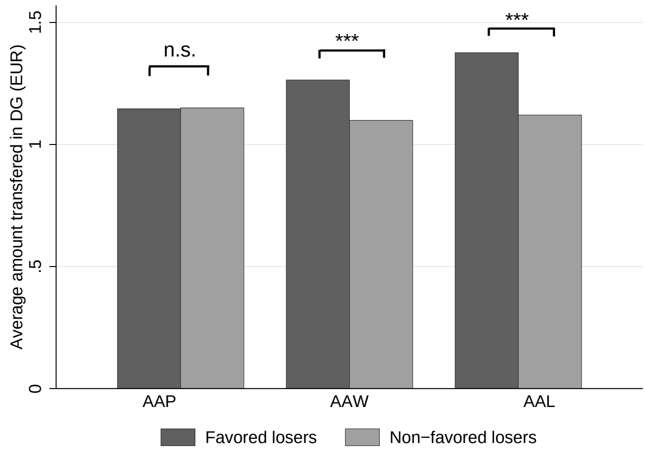

Figure 6. Internalization in dictator games

Notes: Average transfer (EUR) from all group members to favored losers versus non-favored losers in dictator games for each treatment. The brackets and stars above each bar show results of Wilcoxon signed rank tests, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Overall, on average subjects seem to internalize the normative content of those affirmative action policies that are perceived as especially fair and try to act accordingly even beyond the context in which they are binding.

4 Conclusion

One defining feature of all affirmative action policies is to base hiring, promotion, admission decisions etc. not on observed performance alone but to complement or adjust observed performance by further, politically desired criteria. For a given technology, ability, effort and luck are the three key determinants of performance (Cappelen et al., 2007, 2010). In this sense, we investigate the "whole universe" of possible quota rules by analyzing the implications of three different quota rules that favor individuals who score low on one of the three determinants of performance. In particular, the quotas favor individuals with low ability (measured by low productivity), with low effort (measured by choosing a short working time), or enduring bad luck, which resembles discrimination, respectively. One advantage of this stylized approach to study the implications of affirmative action policies is that the motivation to favor certain individuals is clearly stated – quite in contrast to a gender quota, for example, that some will attribute to lower skill levels of women, others to offsetting disadvantages due to part-time work, and still others to unjustified discrimination against women. Explicitly stating the reasons for a favored treatment provides a homogeneous perception of what an affirmative action policy is about and a sound basis for eliciting fairness judgments of such policies that can be attributed to the respective criterion for favored treatment.

While Balafoutas et al. (2016), Calsamiglia et al. (2013), Fallucchi and Quercia (2018), and Petters and Schroeder (2019) study affirmative action policies aiming to compensate for bad luck or discrimination, we are not aware of other attempts to investigate affirmative action favoring individuals with low productivity or short working time, although such policies exist. For example, all countries of the European Union provide part-time employees equal access to promotions even if their overall performance is lower since they work shorter hours. Examples for affirmative action in favor of individuals with low productivity are instances in which individuals with dyslexia or physical restraints get extra time in exams.

Our results document that quotas for discriminated individuals and those who have chosen to work shorter meet their main aim: they effectively encourage the targeted individuals to enter competition. In contrast, a quota for low productivity individuals does not have such an encouragement effect. Compared to a situation without affirmative action, none of the three affirmative action policies under study harms overall efficiency, post-competition teamwork or induces significant retaliation towards the group of favored individuals as a whole. Thus, our results largely reinforce the rather positive findings regarding the consequences of affirmative action policies in studies on gender quotas (e.g., Balafoutas and Sutter, 2012; Balafoutas et al., 2016; Ibanez and Riener, 2018; Kölle, 2017; Niederle et al., 2013) or caste membership (Banerjee et al., 2018, 2019) and extend them to affirmative action policies targeting discriminated individuals and those who perform lower since they have chosen to work shorter hours.

A further result is the high acceptance of affirmative action policies based on judgments regarding their fairness. In particular, affirmative action policies targeting discriminated individuals or those choosing a short working time are judged as significantly fairer than no affirmative action. In times of heated debates about affirmative action, this is important news as it indicates that quotas can get broad political support if they target discriminated individuals or part-time workers and are communicated as such, an aspect that our design ensures.

Perhaps most importantly, our findings suggest that the perceived fairness of affirmative action policies impacts their consequences. Notably, higher fairness perceptions can encourage willingness to compete and prevent retaliation against targeted winners. Additionally, individuals internalize the norms embodied by those affirmative action policies that are rated as fairest and support previously targeted, but unsuccessful individuals even in unrelated post-competition interactions, in which the policy is no longer binding. Affirmative action policies that are perceived as fair thus amplify their impact by influencing behavior beyond the context in which they are applied. From a more general perspective, this finding suggests that not only do preferences shape institutions but that institutions can also shape preferences.

As a whole, our results point at a so far disregarded, but vital aspect in designing and communicating affirmative action schemes, namely their perceived fairness.

Providing a convincing rationale for the implementation of quota rules to ensure that they are perceived as fair seems key to make them a success.

References

Abeler, Johannes, Falk, Armin, Goette, Lorenz, and Huffman, David (2011). "Reference Points and Effort Provision." *American Economic Review*, 101(2), 470–492.

Almås, Ingvild, Cappelen, Alexander, Salvanes, Kjell G, Sørensen, Erik, and Tungodden, Bertil (2016). "Willingness to Compete: Family Matters." *Management Science*, 62(8), 2149–2162.

Araujo, Felipe A., Carbone, Erin, Conell-Price, Lynn, Dunietz, Marli W., Jaroszewicz, Ania, Landsman, Rachel, Lamé, Diego, Vesterlund, Lise, Wang, Stephanie W., and Wilson, Alistair J. (2016). "The slider task: an example of restricted inference on incentive effects." *Journal of the Economic Science Association*, 2(1), 1–12.

Bagde, Surendrakumar, Epple, Dennis, and Taylor, Lowell (2016). "Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India." *American Economic Review*, 106(6), 1495–1521.

Balafoutas, Loukas, Davis, Brent J., and Sutter, Matthias (2016). "Affirmative action or just discrimination? A study on the endogenous emergence of quotas." *Journal of Economic Behavior and Organization*, 127, 87–98.

Balafoutas, Loukas and Sutter, Matthias (2012). "Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory." *Science*, 335(6068), 579–582.

Banerjee, Ritwik, Gupta, Nabanita Datta, and Villeval, Marie Claire (2018). "The spillover effects of affirmative action on competitiveness and unethical behavior." *European Economic Review*, 101(C), 567–604.

Banerjee, Ritwik, Gupta, Nabanita Datta, and Villeval, Marie Claire (2019). "Self Confidence Spillovers and Motivated Beliefs." *GATE Working Paper*, 6.

Beaman, Lori, Chattopadhyay, Raghabendra, Duflo, Esther, Pande, Rohini, and Topalova, Petia (2009). "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics*, 124(4), 1497–1540.

Beaurain, Guillaume and Masclet, David (2016). "Does affirmative action reduce gender discrimination and enhance efficiency? New experimental evidence." *European Economic Review*, 90, 350–362.

Bock, Olaf, Baetge, Ingmar, and Nicklisch, Andreas (2014). "hroot: Hamburg Registration and Organization Online Tool." *European Economic Review*, 71(C), 117–120.

Buser, Thomas, Niederle, Muriel, and Oosterbeek, Hessel (2014). " Gender, Competitiveness, and Career Choices." *The Quarterly Journal of Economics*, 129(3), 1409–1447.

Calsamiglia, Caterina, Franke, Jörg, and Rey-Biel, Pedro (2013). "The incentive effects of affirmative action in a real-effort tournament." *Journal of Public Economics*, 98, 15–31.

Cappelen, Alexander W., Hole, Astri Drange, Sørensen, Erik Ø, and Tungodden, Bertil (2007). "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review*, 97(3), 818–827.

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

Corgnet, Brice, Hernán-González, Roberto, and Schniter, Eric (2015). "Why real leisure really matters: incentive effects on real effort in the laboratory." *Experimental Economics*, 18(2), 284–301.

Council of European Union (1997). "Council Directive 97/81/EC." (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31997L0081).

Croson, Rachel and Gneezy, Uri (2009). "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2), 448–474.

Disability Rights Commission (2007). "Code of practice (revised) for providers of post-16 education and related services."

Falk, Armin, Becker, Anke, Dohmen, Thomas, Enke, Benjamin, Huffman, David, and Sunde, Uwe (2018). "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics*, 133(4), 1645–1692.

Fallucchi, Francesco and Quercia, Simone (2018). "Affirmative action and retaliation in experimental contests." *Journal of Economic Behavior and Organization*, 156, 23–40.

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

Fish, Stanley (2000). "The Nifty Nine Arguments against Affirmative Action in Higher Education." *The Journal of Blacks in Higher Education*, (27), 79–81.

Frederick, Shane (2005). "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives*, 19(4), 25–42.

Fullinwider, Robert (2011). "Affirmative action." In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (editor).

Gächter, Simon, Huang, Lingbo, and Sefton, Martin (2016). "Combining "real effort" with induced effort costs: the ball-catching task." *Experimental Economics*, 19(4), 687–712.

Gill, David and Prowse, Victoria (2012). "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review*, 102(1), 469–503.

Gneezy, Uri, Niederle, Muriel, and Rustichini, Aldo (2003). "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics*, 118(3), 1049–1074.

Goerg, Sebastian J., Kube, Sebastian, and Radbruch, Jonas (2019). "The effectiveness of incentive schemes in the presence of implicit effort costs." *Management Science*, 65(9), 3949–4450.

Greiner, Ben (2004). "The Online Recruitment System ORSEE - A Guide for the Organization of Experiments in Economics." Papers on Strategic Interaction. Max Planck Institute of Economics, Strategic Interaction Group.

Heilman, Madeline E., Block, Caryn J., and Stathatos, Peter (1997). "The Affirmative Action Stigma of Incompetence: Effects of Performance Information Ambiguity." *The Academy of Management Journal*, 40(3), 603–625.

Ibanez, Marcela and Riener, Gerhard (2018). "Sorting through Affirmative Action: Three Field Experiments in Colombia." *Journal of Labor Economics*, 36(2), 437–478.

Ip, Edwin, Leibbrandt, Andreas, and Vecci, Joseph (2020). "How Do Gender Quotas Affect Workplace Relationships? Complementary Evidence from a Representative Survey and Labor Market Experiments." *Management Science*, 66(2), 503–1004.

Jensenius, Francesca Refsum (2015). "Development from Representation? A Study of Quotas for the Scheduled Castes in India." *American Economic Journal: Applied Economics*, 7(3), 196–220.

Kölle, Felix (2017). "Affirmative action, cooperation, and the willingness to work in teams." *Journal of Economic Psychology*, 62, 50–62.

Leibbrandt, Andreas and List, John A (2018). "Do Equal Employment Opportunity Statements Backfire? Evidence From A Natural Field Experiment On Job-Entry Decisions." *NBER Working Paper*, (25035).

Leibbrandt, Andreas, Wang, Liang Choon, and Foo, Cordelia (2017). "Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women." *Management Science*, 64(8), 3501–3516.

Leonard, Kenneth, List, John, and Gneezy, Uri (2009). "Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society." *Econometrica*, 77, 1637–1664.

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

Niederle, Muriel and Vesterlund, Lise (2007). “Do Women Shy Away From Competition? Do Men Compete Too Much?” *The Quarterly Journal of Economics*, 122(3), 1067–1101.

Niederle, Muriel and Vesterlund, Lise (2011). “Gender and Competition.” *Annual Review of Economics*, 3, 601–630.

Petters, Lea and Schroeder, Marina (2019). “Negative Side Effects of Affirmative Action. How Quotas Lead to Distortions in Performance Evaluation.” *CGS working paper*, 8(1).

Raven, John and Raven, C.Jean (2008). “Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven’s Quest for Non-Arbitrary Metrics.” In. Competency Motivation Project. Chap. General Introduction and Overview: The Raven Progressive Matrices Tests: Their Theoretical Basis and Measurement Model.

Rotter, Julian B (1966). “Generalized expectancies for internal versus external control of reinforcement.” *Psychological Monographs: General and Applied*, 80(1), 1–28.

Schotter, Andrew and Weigelt, Keith (1992). “Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results.” *The Quarterly Journal of Economics*, 107(2), 511–539.

Schupp, Jürgen and Gerlitz, Jean-Yves (2008). “BFI-S: Big Five Inventory-SOEP.” *In Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version.* 12.

Sutter, Matthias and Glätzle-Rützler, Daniela (2015). “Gender Differences in the Willingness to Compete Emerge Early in Life and Persist.” *Management Science*, 61(10), 2339–23354.

Sutter, Matthias, Glätzle-Rützler, Daniela, Balafoutas, Loukas, and Czermak, Simon (2016). “Cancelling out early age gender differences in competition: an analysis of policy interventions.” *Experimental Economics*, 19(2), 412–432.

Wagner, Gert G, Frick, Joachim R, and Schupp, Jürgen (2007). “The German Socio-Economic Panel Study (SOEP) - evolution, scope and enhancements.” *SOEPpapers on Multidisciplinary Panel Data Research*, (1).

A Additional results

A.1 Variation in productivity and working time

Figure A.1 and Figure A.2 show the distributions of productivity and working time for all 463 subjects participating in the preparatory session. The median productivity is 7 in the DICE Lab sample and 8 in the BonnEconLab sample. The median working time is 17 min 46 s in the DICE Lab sample and 17 min 17 s in the BonnEconLab sample. In Figure A.2, the spike at 61 minutes is due to the fact that we stopped subjects who still worked on the grid task after 60 minutes. Those who work longer are classified as subjects with long working time anyway and any further measurement of their chosen working time is not necessary.

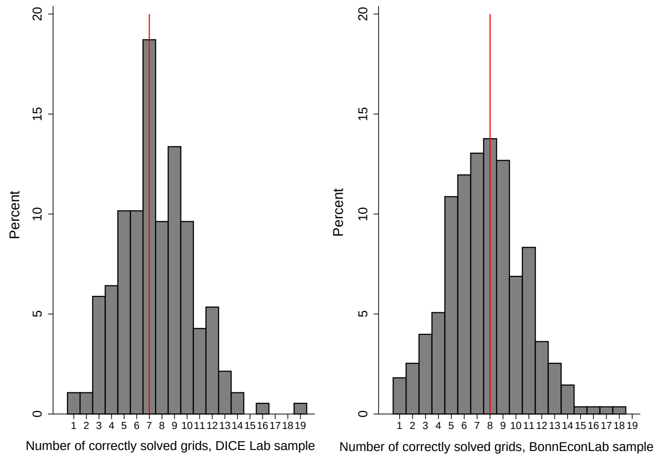

Figure A.1. Distribution and median split (indicated by the vertical red line) of productivity by sample

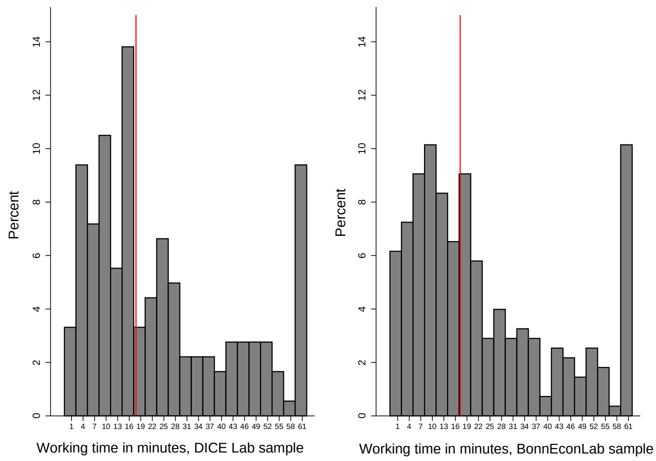

Figure A.2. Distribution and median split (indicated by the vertical red line) of working time by sample

A.2 Fairness perception

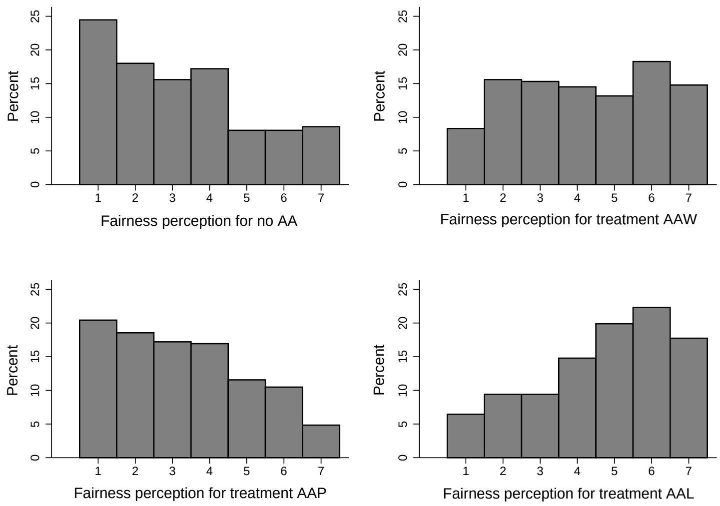

Figure A.3. Distribution of fairness perception

Notes: Fairness perception for no affirmative action and for each affirmative action policy. Higher numbers indicate that a policy is perceived as fairer.

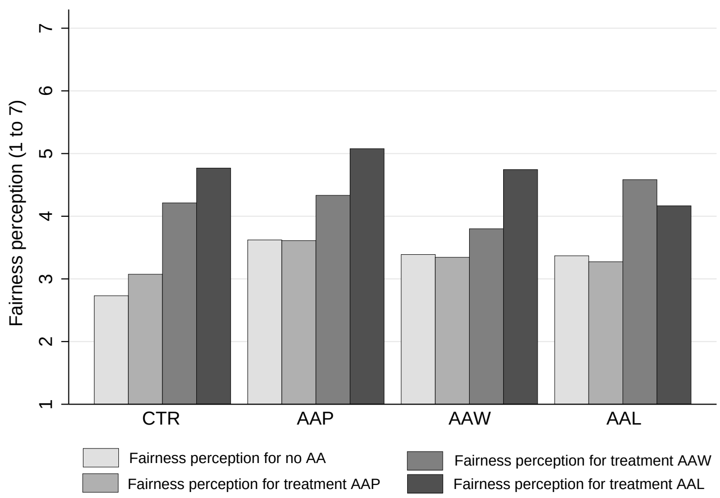

Figure A.4. Fairness perception by treatment

Notes: Fairness perception for no affirmative action and for each affirmative action policy by treatment. CTR/AAP/AAW/AAL on the horizontal axis indicate answers from subjects in treatment CTR/AAP/AAW/AAL. Higher numbers indicate that a policy is perceived as fairer.

A.3 Willingness to compete

	(1)	(2)	(3)
Belief of rank			
2	-0.085 (0.082)	-0.057 (0.096)	-0.104 (0.086)
3	-0.318*** (0.088)	-0.252** (0.104)	-0.336*** (0.095)
4	-0.318*** (0.089)	-0.271*** (0.100)	-0.341*** (0.092)
5	-0.396*** (0.095)	-0.367*** (0.108)	-0.494*** (0.096)
6	-0.507*** (0.098)	-0.406*** (0.118)	-0.564*** (0.098)
Risk attitude	0.028*** (0.010)	0.028** (0.012)	0.031*** (0.012)
Female	-0.095* (0.048)	-0.109* (0.057)	-0.097* (0.055)
Fairness perception	0.016 (0.013)	0.028* (0.015)	0.049** (0.021)
Favored			0.496*** (0.130)
Fairness perception × Favored			-0.069** (0.030)
N	372	264	264

Notes: Average marginal effects from a probit regression. Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The binary dependent variable is willingness to compete in stage 3 (1 if tournament, 0 if piece-rate). *Belief on rank* is a categorical variable about beliefs regarding own rank (between 1 and 6) in stage 3; *Risk attitude* is the answer to the general risk question elicited on an 11-point scale, higher numbers indicate a higher willingness to take risks; *Female* is an indicator variable for gender (1 if female, 0 if male); *Fairness perception* reflects fairness rating of own treatment, elicited on a 7-point scale on which higher numbers indicate higher perceived fairness; *Favored* is an indicator variable (1 if favored type, 0 otherwise).

Table A.1. Individual level determinants of willingness to compete

A.4 Efficiency decomposition

In this section we take a closer look at the different aspects that enter into the efficiency comparison we conduct in Section 3.3.
The measure used there is tournament winners' number of correctly solved grids per minute. This measure actually captures two separate aspects. The first is winner composition which is determined both by the quota rule (or its absence) and

participants' entry decision. The second is how well winners perform under the given incentives.

One thing to keep in mind is that winner selection is not necessarily efficient in the absence of affirmative action. In CTR, the two highest-performing subjects are chosen as winners. Luck and working time are key components of that performance. In fact in CTR in stage 2, only for 28% of groups the winners are the two most efficient subjects, i.e., those who solved most grids per minute correctly in the preparatory session.[26] Theoretically, the quota rule is likely to replace to replace a less efficient individual by a more efficient one as a winner in treatments AAL and AAW, while in AAP this is the other way around. With 36%, 27% and 0% of groups having two most efficient winners in stage 2 in AAL, AAW and AAP, respectively, our data is consistent with this.

We already analyzed the effects of affirmative action on competition entry in Section 3.2.2. Although overall effects were modest, we should take a closer look at whether affirmative action encourages tournament entry in stage 3 of the "right" subjects. For this we look at subjects who are among the two most efficient in their respective group according to the number of correctly solved grids in the preparatory session. While in the absence of affirmative action (CTR), 53% of those most efficient subjects enter the tournament, this is the case for 67%, 44%, and 41% of "most efficient" subjects in AAL, AAW, and AAP, respectively. In consequence, 61% of stage 3 tournament winners in CTR are actually the most efficient in their respective group. This share increases to 69% and 63% in AAL and AAW, and decreases to 38% in AAP.

The efficiency measure reported in Figure 4 accounts for a possible effect of quota rules on motivation in addition to the effects stated in this subsection.

[26] In this section we regard the two subjects who have the highest number of correctly solved grids per minute in the ability task in the preparatory session as the "most efficient". We use the data from that task since in the preparatory session monetary incentives are the same for everyone. In stage 1 of the main session the assigned type might already affect the number of correctly solved grids per minute through monetary incentives and motivation.

A.5 Post-competition measures

	Overall	Productivity type		Working time type		Luck type	
		High	Low	Long	Short	Lucky	Unlucky
CTR	7.20	7.39	7.02	7.31	7.09	7.24	7.17
AAP	7.08 (0.960)	7.56 (0.586)	6.60 (0.652)				
AAW	6.91 (0.481)			7.21 (0.929)	6.61 (0.255)		
AAL	7.35 (0.115)					7.17 (0.483)	7.52 (0.124)

Notes: Average performance (number of correctly solved slider screens) in stage 4. The first column is average performance by treatment. The following columns show average performance of a specific type in the treatment affecting that type and the control treatment. P-values for Mann-Whitney U tests comparing AA treatments with CTR in parentheses.

Table A.2. Average performance in the slider task

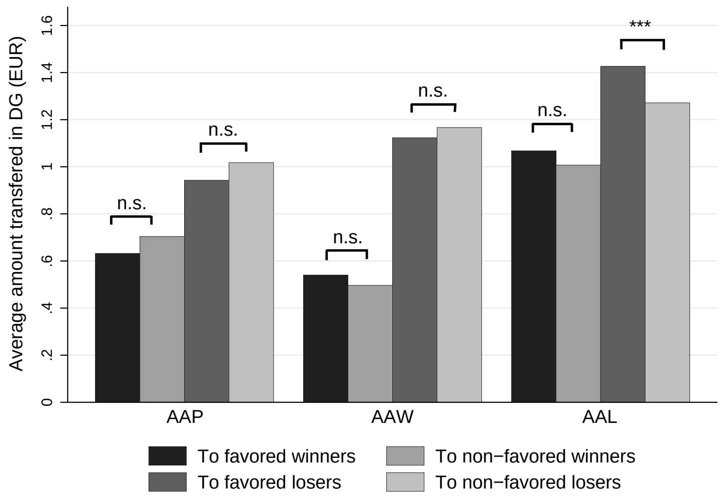

Figure A.5. Retaliation in dictator games - a broader view

Notes: The figure displays the average transfer amount in EUR from non-favored subjects to favored winners versus non-favored winners as well as to favored losers versus non-favored losers in the dictator games for each AA treatment. The brackets and stars above each bar show results of Wilcoxon signed rank tests: * $p < 0.1$; ** $p < 0.05$, *** $p < 0.01$.

B Measurement of fairness perception (for online publication)

The following text is translated from German and refers to the control treatment (CTR). In all treatments, subjects were first asked how fair they perceive the rules of competition they had actually been exposed to to be, before the other competition rules were described and rated.

In the following, we would like to know how fair you perceived the rules of competition in PART 2 of the experiment to be.

As a reminder, in a group of six members, the two members with the highest overall performance (that is, the number of correct answers × the multiplier) were the two winners of the competition.

How fair do you perceive the rules of the competition to be? The leftmost box means "completely unfair", the rightmost means "completely fair". With the boxes in between you can graduate your statement.

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

Now we would like to know, how fair you perceive other possible rules for the competition to be. Just like in the competition you have participated in in PART 2, the following applies in all competition rules: in a group of 6 members, there are two winners who earn a positive amount of money. The other group members earn nothing. With regard to productivity, working time, and multiplier, the group composition is exactly the same as described on page 2 of the instructions.

Each form of competition has one additional special rule:

- Special rule A: At least one winner must be a group member whose productivity lies in the lower half in comparison to all other participants in the experiment. Productivity is defined as the number of correctly solved tasks per minute of a participant in the five-minute counting task last week.
- Special rule B: At least one winner must be a group member whose working time last week was in the lower half in comparison to all other participants in the experiment, and who therefore worked for 4.5 minutes on the task today.
- Special rule C: At least one winner must be a group member with the low multiplier of 0.75.

How fair do you perceive the competition with special rule A? The leftmost box means "completely unfair", the rightmost means "completely fair". With the boxes in between you can graduate your statement.

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

How fair do you perceive the competition with special rule B?

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

How fair do you perceive the competition with special rule C?

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

C Experimental instructions (for online publication)

C.1 Preparatory session

The preparatory session started with an on-screen description of the grid task and an unpaid trial round of the grid task (including feedback on whether each table was solved correctly). We then measured baseline productivity, implemented the questionnaire (for details, see Appendix D), and finally measured the individual choice of working time. Below we provide translated versions of the instructions that were originally in German.

Measurement of baseline productivity: On-screen instructions for the five-minute grid task
You will now start working on the task. Your performance in this task is relevant for your payment. The more counting tasks you solve correctly in the given time, the higher your payment. For each correctly solved table, you receive 0.50 EUR. In the upper right corner of the screen, the remaining time (in seconds) is shown. The task lasts 5 minutes.
Please try hard to solve as many tables correctly as possible in the five minutes, so that we get a realistic idea how good you are in this task.

Measurement of choice of working time: On-screen instructions for the grid task in which subjects choose their working time
You will now again work on a similar task as the previous one. As before, you count the number of zeros ("0") in each table and receive 10 Euro-cent for it, i.e. 10 Euro-cent for each correct table. However, you are now free to choose how long you like to work on the task. You will start working on the task on the next screen, and can work on it as long as you want. The tables will appear one after another, until you decide to stop working. In addition, there is a special feature: Your working time today will also determine your working time in the experiment next week. Next week, you will work on a similar task again in which you will be given a specific amount of time to solve as many tables as possible and get paid accordingly. In this task, the tables will appear one after another until working time is up. Based on your working time today, we will form two groups of the same size but with different working times. Those who decide to work shorter today, will also work shorter next week (one half of the participants). Those who decide to work longer today, will also work longer next week (the other half of the participants). Individuals who work shorter will, on average, solve fewer tables correctly and therefore earn less. For them, the experiment will be shorter (it will end earlier). Individuals who work longer will, on average, solve more tables correctly and therefore earn more. For them, the experiment will be longer (it will end later). In case you have any questions, please raise your hand. An experimenter will then come to your seat.

C.2 Main session

General instructions (distributed on paper at the beginning of the main session)

Welcome to today's experiment! Thank you for participating!

During the experiment, you and the other participants will be asked to make decisions. Your own decisions as well as the decisions of the other participants will determine your earnings, according to the rules that will be described in what follows.

The experiment will be conducted on the computer. You make your decisions on the screen. All your decisions and answers will remain confidential and anonymous.

The experiment consists of 5 parts. PART 1, PART 2, PART 3, PART 4, and PART 5. Additionally, you will answer a short questionnaire.

One of the five parts will be selected randomly by the computer to determine your payment. Every part of the experiment is equally likely to be selected. It is therefore in your own interest to make your decisions in each part as if it was the only part.

Independent of your decisions you will receive a show-up fee of 4 EUR. This means that your total earnings from today's session will be the payment from the randomly chosen one of the five parts of the experiment plus the show-up fee of 4 EUR. You will receive your earnings at the end of today's session together with the earnings from last week.

All other explanations will be given stepwise at the beginning of each part of the experiment. You will receive the instructions for each part in turn. You will have enough time to read the instructions carefully and to ask questions. Please do not hesitate to ask questions if something is unclear.

Please note that, as the last week, talking is not permitted. If you have questions, please do not ask them loudly but raise your hand. One of the experimenters will come to your seat to answer your question. If you do not comply with these rules you will be excluded from the experiment and you will not receive any payments.

General information regarding today's experiment

In today's experiment, your task is once again to solve as many counting tasks correctly as possible in a given amount of time, i.e. to correctly count the number of zeros (“0”) in as many tables as possible. In addition, there is one special feature. **Each participant has three characteristics which remain fixed during the whole experiment: his productivity, his working-time (in minutes) and his multiplier**.

- The **productivity** states how many counting tasks per minute the participant has solved correctly in last week's five-minutes-task. For half of the participants, productivity lies **in the lower half**. For the other half of participants, productivity lies **in the upper half**.

- Today's **working time** depends on the self-chosen working time in the task at the end of the session last week. Half of the participants will have **4.5 minutes** per task today to solve as many tables correctly as possible. These are those participants whose working time belonged to the lower half last week. The other half of the participants will have **7.5 minutes** per task today to solve as many tables correctly as possible. These are those participants whose working time belonged to the upper half last week.

- The **multiplier** is a number which is multiplied with the number of correctly solved counting tasks to determine overall performance. The multiplier will be assigned randomly to each participant. For half of the participants, the multiplier will be **0.75**. For the other half, it will be **1.25**.

You will soon receive information about your productivity, your working time and your multiplier on the following screen.

The performance of each participant is determined as follows:

Performance = Number of correctly solved counting tasks in your working time × Multiplier

PART 1 – Piece rate (distributed on paper at the beginning of stage 1)

Your task in PART 1 is similar to the one in the first session. Again, the task is to solve as many counting tasks as possible in a given amount of time, i.e. to correctly count the number of zeros (“0”) in as many tables as possible. How much time you have is displayed on the screen. Each table consists of ten rows and ten columns, which contain either a zero (“0”) or a one (“1”). Each table differs from the previous one. You are allowed to use the provided scratch paper if you like. After you have entered your response, please click the “confirm” button. Afterwards, you will learn immediately on the same screen whether your answer is right or wrong.

If PART 1 of the experiment is chosen for payment, you will receive the following payment:

$$\textbf{Payment} = \underbrace{\textbf{No. of correctly solved counting tasks in your working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{0.50 EUR}$$

For example, if you have solved **ten** tables correctly and your multiplier is **1.25**, you receive the following payment:

Payment = 10×1.25×0.50 EUR = 6.25 EUR

If you have answered **ten** questions correctly and your multiplier is **0.75**, you receive the following payment:

Payment = 10×0.75×0.50 EUR = 3.75 EUR

Your payment will not be reduced if you enter a wrong answer. We will refer to this payment as the **piece-rate payment** from now on.

After all questions regarding PART 1 are answered, your working time for PART 1 will start.

PART 2 – Tournament (distributed on paper at the beginning of stage 2)

As in PART 1, you will have a given amount of time to solve as many counting tasks correctly as possible. Again, your working time is displayed on the screen. Different from before, in this part your payment depends on your performance relative to the performance of other participants in your group.

Group allocation:

For the following parts of the experiment, you will be allocated to a **group with 6 members**. The groups were formed randomly and stay the same throughout the whole experiment. This means that you will form a group with the same participants for the rest of the experiment.

Reminder: **Each participant has 3 characteristics: his productivity, his working time, and his multiplier.**

Note that each group consisting of six **members** meets the following criteria regarding productivity, working time and multiplier:

- The **productivity** of **three** group members lies in the **upper half** compared to all participants. The productivity of the **other three** group members lies in the **lower half** compared to all participants.
- The chosen **working time** last week of **three** group members lies in the **upper half** compared to all participants. Therefore, these three group members work for **7.5 minutes** on each counting task today. The chosen **working time** last week of the **other three** group members lies in the **lower half** compared to all participants. Therefore, these three group members work for **4.5 minutes** on each counting task today.

- The randomly drawn **multiplier** of **three** group members is **0.75**. The number of correctly solved tables of these three group members will thus be multiplied with **0.75** to calculate overall performance. The randomly drawn **multiplier** of **the other three** group members is **1.25**. The number of correctly solved tables of these three group members will thus be multiplied with 1.25 to calculate overall performance.

Rules of the tournament:

If PART 2 is chosen for payment, your payment depends on how high your performance is compared to the other five members of your group.

The **two** group members with **highest** overall performance (i.e. number of correctly solved tasks in the total individual working time × Multiplier) are the two winners of the tournament.

(The content of the following part in gray differs across treatments. There is no further content for the control treatment (CTR).)

Affirmative action w.r.t. productivity (AAP):

In addition, the following **special rule** is applied:

At least one winner must be a group member whose productivity lies in the lower half in comparison to all other participants in the experiment. Productivity is the number of correctly solved counting tasks per minute last week.

If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose productivity lies in the lower half will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**

Affirmative action w.r.t. working time (AAW):

In addition, the following **special rule** is applied:

At least one winner must be a group member whose working time last week lied in the lower half in comparison to all other participants in the experiment, and who therefore works for 4.5 minutes on the task today.

If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose working time is 4.5 minutes will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members is no longer a winner.**

Affirmative action w.r.t. luck (AAL):

In addition, the following **special rule** is applied:

At least one winner must be a group member with the low multiplier of 0.75.

If this is not automatically the case given the overall performance of the group members, then the group member with the highest performance of the three group members with the low multiplier of 0.75 will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members is no longer a winner.**

The payment of the two winners is as follows:

$$\textbf{Payment} = \underbrace{\textbf{Correctly solved counting tasks in their working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{1.50 EUR}$$

For example, a winner with 10 correct answers and a multiplier of **1.25** receives the following payment:

$$\textbf{Payment = 10} \times \textbf{1.25} \times \textbf{1.50 EUR = 18.75 EUR}$$

A winner with 10 correct answers and a multiplier of **0.75** receives the following payment:

$$\textbf{Payment = 10} \times \textbf{0.75} \times \textbf{1.50 EUR = 11.25 EUR}$$

The **other four** members of your group get **no payment**.

If there is a tie between two group members, the winner will be determined randomly. We will refer to this payment as **tournament payment** from now on. At the end of today's session, you will be informed about the outcome of the tournament.

PART 3 – Choice between piece-rate and tournament payment

(distributed on paper at the beginning of stage 3)

Similar to PART 1 and PART 2, you will have a given amount of time to solve as many counting tasks correctly as possible. Your working time will be shown on the screen.

However, now you choose by yourself which payment scheme you prefer for your performance in PART 3. You can choose either the piece-rate payment (same rules as in PART 1) or the tournament payment (same rules as in PART 2).
If PART 3 is chosen for payment, your earnings will be determined as follows:

- If you choose the **piece-rate payment**, your payment is:

$$\textbf{Payment} = \underbrace{\textbf{No. of correctly solved counting tasks in your working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{0.50 EUR}$$

- If you choose the **tournament payment, your earnings depend on the level of your overall performance in PART 3 compared to the overall performance of your five group members in PART 2 (tournament).** Reminder: PART 2 is the part you have just finished.

(The content of the following part in gray differs across treatments.)

Control treatment (CTR):

If your overall performance (i.e. number of correctly solved counting tasks in the individual working time × Multiplier) is higher than that of at least four other members of your group in PART 2, your payment is as follows:

Affirmative action w.r.t. productivity (AAP):

In general, the two group members with the highest overall performance, i.e. (number of correct answers in the total individual working time) × (Multiplier), are the two winners of the competition.

The following **special rule** is still applied:

At least one winner must be a group member whose productivity lies in the lower half in comparison to all other participants in the experiment. Productivity is the number of correctly solved counting tasks per minute last week.

If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose productivity lies in the lower half compared to all participants in the experiment will **replace** the initial second-best winner.

In this case the group member with the second highest performance of all six group members of your group is no longer a winner.

If your overall performance in PART 3 relative to the overall performance of your group members in PART 2 implies you are a winner, your payment is as follows:

Affirmative action w.r.t. working time (AAW):

In general, the two group members with the highest overall performance, i.e. (number of correct answers in the total individual working time) × (Multiplier), are the two winners of the competition.

The following **special rule** is still applied:

At least one winner must be a group member whose working time last week lied in the lower half in comparison to all other participants in the experiment, and who therefore works for 4.5 minutes on the task today.

If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose working time is **4.5 minutes** will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members is no longer a winner.**

If your overall performance in PART 3 relative to the overall performance of your group members in PART 2 implies you are a winner, your payment is as follows:

Affirmative action w.r.t. luck (AAL):

The following **special rule** still is applied:

At least one winner must be a group member with the low multiplier of 0.75.

If this is not automatically the case given the overall performance of the group members, then the group member with the highest performance of the three group members with the low multiplier of 0.75 will **replace** the initial second-best winner. **In this case the group member with the second**

highest performance of all six group members is no longer a winner.

If your overall performance in PART 3 relative to the overall performance of your group members in PART 2 implies you are a winner, your payment is as follows:

$$\textbf{Payment} = \underbrace{\textbf{No. of correctly solved counting tasks in your working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{1.50 EUR}$$

That means it is three times as high as the piece-rate payment.

If your overall performance in PART 3 relative to the overall performance of the other group members in PART 2 implies that **you are not a winner**, you get **no payment**.

If there is a tie between two group members, the winner will be randomly determined.

The group composition is the same as in PART 2. If you choose the tournament payment, you will be informed about the outcome of the tournament at the end of the experiment.

On the next screen, you will decide whether you choose the piece-rate payment or the tournament payment for your performance in PART 3. Then the task will begin.

PART 4 (displayed on screen at the beginning of stage 4)

In the following, you will work on a new task in which you have to place slider markers in a certain position.

You will see six sliders on each screen. They can be placed on a scale from 0 to 100. As soon as you click on a slider marker, the current position will be displayed on the screen. You can change the position using the mouse.

Your task is to move all six slider markers on a screen to the position of "50". Only then a screen is finished correctly and you can proceed to the next screen by clicking the "Continue" button. You have five minutes to correctly finish as many screens as possible. In this task, all participants work for the same amount of time and there is no multiplier.

Your payment in this part depends on the number of screens that you and the other five members of your group finish correctly. The group composition is the same as before.

Precisely, your payment is determined as follows: You will receive 10 Euro-cent for each correctly finished screen by each member of your group (including yourself). The other members of your group will also receive 10 Euro-cent for each screen that any group member (including yourself) has finished correctly. This means each correctly finished screen by each player yields 60 Euro-cent for the group (i.e. all six group members together).

In addition, the members of the group who won the tournament in PART 2, will receive an endowment of 5 EUR. The other members will receive an endowment of 2 EUR.

If PART 4 is chosen for payment, your payment is the sum of your individual endowment and your earnings from the sum of all correctly finished screens of your group members.

At the end of the experiment, you will be informed about the performance of your group.

If you have any questions, please raise your hand.

PART 5 (displayed on screen at the beginning of stage 5)

In this part, you are asked to make five decisions which will affect you and one of the five other members of your group, respectively. In order to be able to attribute decisions, each group member will be randomly assigned a number from 1 to 6. You are group member number X.[27]

For each decision, you will get an initial endowment of 5 EUR. Your task is to decide how to split this endowment between you and the other member of your group. You may choose an amount between 0 and 5 EUR (in steps of 10 Euro-cents) which you want to pass on to the other group member. You will keep the rest for yourself. You will not get any information about the identity of the other group member and the other group member will not get any information concerning your identity. The only thing you will get to know about the respective other group member before you will make your decision is whether (s)he has won the tournament in PART 2 or not.

(The content of the following part in gray differs across treatments. There is no further content for the control treatment (CTR).)

[27] The exact number differs for each subject.

Affirmative action w.r.t. productivity (AAP):
... and whether his/her productivity lies in the upper or in the lower half.

Affirmative action w.r.t. working time (AAW):
... and his/her working time.

Affirmative action w.r.t. luck (AAL):
... and his/her multiplier.

If this part is chosen for payment, your payment will be determined as follows: In each group, three pairs are chosen randomly and their decisions will determine payments. This means each group member is assigned to exactly one pair which is relevant for the payment. In each pair, it is randomly assigned who will be the donor and the recipient. The decision of the donor determines the payment of both. This means that each decision is paid out with the same probability and therefore you should make each decision as if it was the only one.

At the end of the experiment, you will be informed about the number of the group member you paired with, who is the donor and who is the recipient in this pair, and what the donor has decided. You will not get any information about the decisions made in the other pairs (that you do not belong to).

If you have any questions, please raise your hand.

D Questionnaire (for online publication)

The questionnaire in the preparatory session contains the following items:

1. **Risk preference, general risk question:** same wording as in German Socio-Economic Panel questionnaire (SOEP, see, for example, Wagner et al. (2007))
How do you evaluate yourself? Are you generally a risk-seeking person or do you try to avoid risks? The leftmost box means "not at all risk-seeking" and the rightmost "very risk-seeking". With the boxes in between, you can graduate your statement.

not at all risk-seeking ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ *very risk-seeking*

2. **Risk preference, incentivized choice list:** Subjects make eleven, pairwise decisions between a lottery with a fifty-fifty chance of winning either 2 EUR or 7 EUR and a safe payment. The safe payment increases in 0.5 EUR increments, ranging from 2 EUR to 7 EUR.

3. **Social preference** (survey question, Falk et al., 2018)
   Question 1: Imagine the following situation: Today you unexpectedly received 1000 EUR. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed).
   Question 2: Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize that you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about 20 EUR in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 EUR, the most expensive one costs 30 EUR. Do you give one of the presents to the stranger as a "thank you" gift?
   Which present do you give to the stranger?
   1. No, would not give present
   2. The present worth 5 EUR
   3. The present worth 10 EUR
   4. The present worth 15 EUR
   5. The present worth 20 EUR
   6. The present worth 25 EUR
   7. The present worth 30 EUR

4. **Big Five:** we use the 15-item Big Five scale developed for the SOEP (Schupp and Gerlitz, 2008) to measure personality traits.

5. **Locus of control:** we use 10 different items adapted from Rotter (1966) which have been used in the 2005 wave of the SOEP.

6. **Questions on general fairness ideals:** all using the same scale

   completely disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely agree*

   To what extent do you personally agree with the following statements?
   It is unfair for someone who does a strenuous job to earn little.
   Who performs better, should earn more.
   If someone is naturally good at something, it is right to reward him/her for it.
   It is wrong to favor somebody just because he/she may have experienced discrimination elsewhere.

7. **Cognitive ability:** Raven matrices from the Wechsler IQ test (J. Raven and C. Raven, 2008).

   Before subjects start the test, we elicit their belief about individual rank as follows:

   Before you begin, we would like to ask you to assess how well you will score in the IQ test compared to the other participants in the experiment. For example, 0-10% means that you are among the 0-10% participants with the fewest correct answers, and at least 90% of the participants have more correct answers than you. 41-50% means that at least 40% of participants have fewer correct answers and at least 50% have more correct answers than you. 91-100% means that at least 90% have fewer correct answers than you. What do you think? How do you compare to the rest of the group?
   ☐ 0-10%
   ☐ 11-20%
   ☐ 21-30%
   ☐ 31-40%
   ☐ 41-50%
   ☐ 51-60%
   ☐ 61-70%
   ☐ 71-80%
   ☐ 81-90%
   ☐ 91-100%

8. **Cognitive reflection test:** see Frederick (2005).

9. **Socio-demographics:** age, gender, final grade point average at academic high school, last math grade at academic high school, field of study, monthly disposable amount of money, political orientation, number of experiments already participated in the same lab.