

Quotas and Cooperation

Johanna Mollerstrom[†]

Job Market Paper

November 2012

Abstract: Selection by quotas is an important policy measure in the affirmative action tool box. However, quotas may come with unintended side effects, for example by causing uncooperative behavior in the group formed with quota-based selection rules. In the laboratory I measure the impact of a quota on group cooperation, and examine the underlying mechanisms. Two groups are created by randomly assigning participants to either an orange or a purple group. In the unrepresentative quota treatment, orange participants are chosen as members of a selected group by performance in a simple unrelated math task whereas purple participants are chosen based solely on the quota. I compare contributions in a public good game in this unrepresentative quota treatment to behavior in a control treatment, where the orange and purple participants are treated symmetrically and all members of the selected group are chosen based on performance on the unrelated math task. My results show significantly less cooperation in the quota treatment and I furthermore find that this tendency is observed in both the meritocratically chosen orange participants and the quota-advantaged purple participants, and regardless of the color of the matched player. The reduced cooperation remains even when participants are given a rationale for the unrepresentative quota, e.g., by appealing to a fairness argument. The negative effect on cooperation from the unrepresentative quota disappears when selection is done completely randomly instead of on the basis of performance.

Key words: discrimination, gender gap, experiment, quota, cooperation, public goods

JEL codes: C91, C92, D03, J15, J16, H41

I am very grateful to the Women and Public Policy Program at the Harvard Kennedy School for financial support and to Iris Bohnet for invaluable support and advice. I also want to thank Philippe Aghion, Alberto Alesina, Anna Dreber Almenberg, Armin Falck, Angela Fonseca-Galvis, Francesca Gino, Ed Glaeser, Erika Hall, Oliver Hart, David Laibson, Geordie McRuer, Al Roth, Ariel Dora Stern, Cass Sunstein, Christoph Wollerheim and seminar participants at Harvard University, Harvard Business School, University of Bonn (CEN) and the Stockholm School of Economics for very valuable comments. I am, of course, fully responsible for any remaining errors. Sara Steinmetz provided excellent assistance conducting the experiments.

[†] Department of Economics, Harvard University, jmollers@fas.harvard.edu.

1. Introduction

Selection by quotas is a commonly used policy tool for getting closer to a desired representation between genders, races or ethnicities. In political settings, the use of quotas has long been widespread (Krook, 2009) and its use has now also reached other spheres, such as the corporate world (see e.g. Ahern and Dittmar, 2012). Even though quotas are often effective in reaching a specific numerical representation, there can be unintended side effects (Pande and Ford, 2011). The focus of this paper is on one such effect, namely that of cooperation problems in groups created by quotas.

In 2003, the Norwegian government unexpectedly announced that the country's public limited- and state-owned companies would be required to have at least 40 percent representation of each gender on their boards by 2008 to avoid forced liquidation. The policy has been found to have affected the management practices of boards and adversely influenced short-run profits (Matsa and Miller, 2011; Ahern and Dittmar, 2012). It has been hypothesized that one change in management practices was related to cooperation problems, which some boards experienced as a consequence of the quota (Stenseng, 2010; Dagens Naeringsliv, 2010; see also Clarke, 2010).¹

Inefficient and uncooperative behavior following quotas or other procedures with preferential selection can also be found in other contexts, for example in education. In the United States, quotas for minority students have been used at all levels of schooling in an attempt to increase diversity of the student body. Sometimes the consequences have been dramatic, as was the case during the Boston riots in the 1970s following the institution of quotas for, and busing of, minority children to previously primarily white schools (Lukas, 1986). At both the college and the graduate school level quotas have been documented to lead to uncooperative behavior, for example in the form of reluctance to share information, both within the student

¹ Huse and his coauthors have written extensively about the experience of women on corporate boards in Norway. In particular the discussion about how board members who enter the board by different procedures, for example women or employee representatives, may experience lower esteem than other board members is relevant for this paper. See e.g. Huse and Solberg (2006), Huse et al. (2009), and Nielsen and Huse (2010). See also Terjesen et al. (2009).

population and between students and teachers (Dreyfuss and Lawrence, 1979; Crosby and VanDeVeer, 2000).

There are also examples of how affirmative action in the workplace can lead to cooperation problems. One area where this has been found is among firefighters. In the United States, affirmative action and quotas have been used since the late 1960s to increase representation of female and minority firefighters. This has been claimed to lead to a decrease in cohesion and cooperation between colleagues, both in training and on the job (Dreyfuss and Lawrence, 1979; Chetkovich, 1997).

The above examples suggest that quotas can give rise to cooperation deficits. But whereas a significant amount of research has been dedicated to impact of quotas on group quality, there has been little systematic research regarding effects on cooperative behavior. The aim of this paper is to investigate whether preferential selection in the form of an unrepresentative quota has a negative impact on group cooperation and, if so, to examine the underlying mechanisms.

The experimental method has some limitations. Key factors from the “real world” settings discussed above, such as interaction over longer time periods, will not be captured in the laboratory setting. However, laboratory experiments also have advantages: the controlled environment in the lab allows us to abstract away from pre-formed groups, such as gender or race, and focus on the process of how a group is formed. The laboratory environment also enables the researcher to use randomization to identify causation.

In the experiment described in this paper groups are created by randomly allocating the 16 participants in each experimental session a color: either orange or purple. Participants are asked to do a math task and in the main two treatments their performance in the math task later serves as a selection device into a selected group. This selected group has eight members and is constrained to have equal representation of purple and orange participants. By varying the color composition of the initial group of 16 participants I create a control treatment where the two colors are treated symmetrically and all participants are chosen by performance in the math task and an unrepresentative quota treatment where the purple

members are admitted into the selected group solely on the basis of the quota, while the orange players are still selected based on performance.

This design captures two aspects of how unrepresentative quotas are often perceived: first, that one group is treated preferentially over the other in the selection process and second, that less weight is given to performance when selecting candidates from this group.²

In order to study cooperative behavior, I next let the eight participants in the selected group play a two-person public good game. This is designed so that all members take part in seven public good games: one against each other member of the selected group. Furthermore, the participants know the color of the person they are playing in each public good game. I find that cooperation is significantly lower when the selected group is created by the unrepresentative quota. In that case, participants contribute about 30 percent of their endowment to the public good, compared to a contribution rate of over 50 percent in the control treatment. I call this decline in cooperation the *quota effect*.

In addition to establishing this result, this design makes it possible to draw some inferences about the mechanism underlying the quota effect. Various theories offering explanations for the quota effect have different predictions about whether the quota effect should be relation-specific (i.e. only be present in some relations in the group, for example when those who were not advantaged play with those who were advantaged) or general (i.e. apply similarly to all relations). I find that the quota effect is general, a finding which contradicts some theories, for example about the quota effect having its roots in a desire by those who are not advantaged by the quota to punish those who are advantaged.

The experimental design also addresses a key policy aspect of quotas, namely the question of justification. Following on work which has shown that the acceptance of affirmative action policies can be affected by whether, and how, the policy is justified (see e.g. Murrell et al., 1994 and Heilman et al., 1996), I introduce two separate justifications for the unrepresenta-

² These are not all potential aspects of unrepresentative quotas (others include slot-taking and fixed representation), but I limit the focus to these two aspects in this experiment. Note also that these two features, (1) preferential selection and (2) different role given to performance may describe also other situations than selection by quotas where favoritism plays a role.

tive quota, the first emphasizing efficiency gains and the second emphasizing fairness. However, even though participants report that they find our fairness argument especially convincing, they do not become more cooperative when this justification is provided.

As discussed above, the unrepresentative quota in this experimental design captures two aspects of how affirmative action policies can be perceived. First, that it is a selection process that gives preferential treatment to one group over another. Second, that performance plays less of a role in selection for candidates from the group that is treated preferentially. In order to understand the respective role of these two aspects in causing the decrease in cooperation in the quota treatment, I also conduct two treatments that are identical to the two main treatments described above with the only difference that selection by performance is replaced with random selection. Hence, in this version of the quota treatment, preference was still given to one group, but the role of performance is the same for both groups. I find that this version of the unrepresentative quota has no negative impact on cooperation relative to the control treatment of symmetric treatment.

The research presented in this paper is related both to the work on procedural fairness and procedural justice. The latter originally aimed at understanding the feelings of entitlement which can arise when individuals earn their roles in an experiment. (Hoffman and Spitzer, 1985). A well-known result in this strand of literature is due to Hoffman et al. (1994, 1996), who show that when the role as first mover in a dictator- or ultimatum game is earned, for example by answering a quiz, offers to receivers are significantly lower than when the role is allocated randomly. In the literature on procedural justice, it has instead been emphasized that a procedure which is perceived as just may provide social motivation and thereby influences cooperative behavior (Lind and Tyler, 1988; Tyler and Lind, 1992; Tyler and De Cremer, 2006).

This paper is also related to the research of Balafoutas and Sutter (2012). Their starting point is the work of Niederle et al. (forthcoming), who show that an affirmative action procedure which guarantees women equal representation among the winners of a competition makes them as likely to enter a competitive setting as equally qualified men (but in the absence of

the affirmative action policy women under-select into the competition, see also Gneezy et al., 2003; and Niederle and Vesterlund, 2007).

Balafoutas and Sutter (2012) conduct an experimental investigation of the effects of such an affirmative action procedure on things other than willingness to compete. They follow the design of Niederle et al. (forthcoming) and add a post-competition teamwork task and a coordination game. They find no effects from the quota on performance in the teamwork or the coordination task. Importantly however, the team task and the coordination game that they use do not incorporate conflicting motives between the group and the individual, i.e. their games are not social dilemmas. Our research, on the other hand, uses a public good game in order to specifically study cooperation.

The rest of the paper is organized as follows. Section 2 describes the experimental design, provides a conceptual framework and generates testable hypotheses. In Section 3, I outline and discuss the results from the experiment. Section 4 discusses different theories that make predictions about the effects of an unrepresentative quota and relates those theories to the experimental results. Section 5 concludes.

2. Experimental Design

In the experiment 16 people participated in each session. After being seated in the laboratory, participants were told that the study would have several parts in which they could earn money. It was also made clear that all the money earned would be paid to them in private at the end of the experiment and that the exchange rate between points in the experiment and dollars was such that 10 points corresponded to 3 dollars.³

After these instructions were explained, the 16 people in each session were randomly allocated a color: either orange or purple.⁴ These colors were chosen since they are relatively neutral for American participants (unlike for example red vs. green). The participants were

³ All experimental instructions can be found in Appendix A.

⁴ The fact that colors were allocated randomly ensured that the characteristics of the two groups were similar, e.g. with regards to gender, age and ethnicity. See Appendix D for details.

told their color allocation on the computer screen and were also given a silicon bracelet with their color to wear for the duration of the experiment. Thereafter they spent two minutes filling out a paper form with five associations to their color. This was done in order to give the participants some time to internalize their randomly allocated color.

After randomization of colors, participants were told that they would be given a number of math tasks where each task would consist of adding up five two-digit numbers and that they would be paid 1 point (30 cents) for each correct answer. They had five minutes available to do as many tasks as possible with the maximum available tasks being 15. It was made clear that the number of tasks that they answered correctly would not be revealed until the end of the experiment.⁵

After having solved math tasks for five minutes, part three of the experiment followed. Participants were told that out of the 16 people in the session, eight would be chosen as members of a selected group called the “high-stake group”. The other eight participants would remain in the study as members of the “regular-stake group”. The instructions made clear that the members of the two groups would do the same thing in the rest of the experiment, but that the high-stake group members would have the chance to earn more money. This design and the naming of the two groups was chosen in order to make it desirable to the participants to be selected to the high-stake group without imposing a specific context on the situation.

It was also explained to the participants that the high-stake group would have four orange and four purple members. By varying the color composition of the underlying group of 16 participants (a composition which was made clear to all participants) this either meant that the two colors were treated symmetrically (the underlying group then consisted of eight purple and eight oranges players) or an unrepresentative quota favoring the purples (in these sessions the underlying group contained four purple and twelve orange players). Before the

⁵ This task has been used previously by among others Niederle and Vesterlund (2007, 2010) and performance has generally been found not to differ between different groups such as between gender or ethnicities. I confirm that there are no differences in performance between people of different ethnicities, but I find both gender and age differences. See Appendix B.

participants were told whether they had been chosen for the high-stake group or not, they were given instructions for the game that was to be played in the next part.

The aim of part four of the experiment was to measure the degree of cooperativeness within the high-stake group. I chose to do this with one of the classic social dilemmas, a two person public good game. In this game, participants were assigned to pairs and given a personal endowment of 20 points (USD 6) in the high-stake group and 10 points (USD 3) in the regular-stake group. I selected a two-person public good game instead of a n -person public good game (with $n > 2$) so that it could be made clear to the participants whom they were playing each game with, in order to be able to assess whether participants treated people of different colors differently.

The public good game was conducted in a standard way (see Ledyard, 1995 for a survey of public good game experiments), and both participants in the pair had to choose how much of their endowment to keep and how much to contribute to a joint project. Points contributed by the two participants were summed together, multiplied by 1.5 and then distributed equally between them. Everyone played the game with each of the other seven members of their group (either the high-stake- or the regular-stake group) and the contribution decisions were made simultaneously in each of the seven games. Participants were told that they would be paid according to the outcome of one randomly chosen game out of the seven that they played. The only thing the participants were told about the person they played with in a specific game was that person's color.

After the participants had made their choices in all games, the experiment proceeded to a questionnaire in which the participants were asked to report their gender, age and ethnicity. In addition, the participants were asked if they found the process by which the members of the high-stake group was chosen fair.

2.1. Selection by performance: Treatments 1 and 2

As described above, the high-stake group was put together such that it would consist of four purple and four orange players. In the main treatments the selection was done so that the four players of each color who performed best in the math task were admitted into the high-stake

group. In the unrepresentative quota treatment (treatment 1), there were twelve orange and four purple players in the underlying group of 16. This implies that when the four best performers of each color were selected, performance did not matter for the purple players, since all of them were automatically selected. This selection process was hence characterized by a quota that preferentially treated the purple players, and made performance in the math task relevant only for the selection of the orange players.

I compare the participants' behavior in the unrepresentative quota treatment (treatment 1) to the participants' behavior in the control treatment (treatment 2). In this treatment there was equal representation of purple and orange players in the original group of 16 – i.e. there were eight orange and eight purple players. This meant that the two colors were treated symmetrically and the eight players who performed best in the math task in part 2 were in expectation selected into the high-stake group.⁶

With the data from treatments 1 and 2 we can compare behavior in the public good game in a high-stake group consisting of four players of each color and which differs only in how they were selected: either by a process characterized by an unrepresentative quota for the purple, or by a process where all eight high-stake group members were selected by performance and symmetrically treated, regardless of color. This comparison allows us to investigate whether cooperative behavior differs depending on the selection process. In addition, by examining if the players' behavior depends on their own color and/or on the color of the person they are playing a particular game with, we can draw some inferences about the mechanism at work.

2.1.1 Justification of the Unrepresentative Quota: Treatments 1b and 1c

In the unrepresentative quota treatment (treatment 1) described above, participants were not provided any explanation for the preferential treatment that the purple players were given. When quotas are used it is most often for one of two reasons. A first argument that is com-

⁶ Note that selection may still not be entirely by performance. If, for example, there are five purple and three orange players among the eight best performers, only the four best of the purples will be selected and one orange player who is not among the eight best will be admitted into the high-stake group. As a future extension of this work, it would be interesting to compare this situation with one where the eight best players are chosen, completely disregarding their colors.

monly used for quotas is that they are there to enhance efficiency. For example, advocates of corporate-board quotas for women claim that they will lead to more valuable perspectives being represented on the board, which in turn will improve the board's decision making and the company's performance (see e.g. Daily and Dalton, 2003; and Huse and Solberg, 2006). Second, it is sometimes argued that a quota corrects unfairness. Examples of this can be found in the policy of affirmative action in the United States that gives preference to members of minority groups (for example African Americans) with reference to this process being a compensation for previous unjust treatment of members of this community (see e.g. Heilman et al., 1996 and Murrell et al. 1994).

To be able to address this practical policy aspect of quotas, I designed two additional treatments that were versions of the unrepresentative quota treatment, in which justifications along the above lines were given.

Treatment 1b was identical to treatment 1 with the exception that an efficiency argument for the quota was given: When the difference between the high-stake and the regular-stake group was introduced to the participants, the instructions explained that the payoff rule was such that the members of the high-stake group would only have a higher endowment than the members of the regular-stake group if the high-stake group had an equal proportion of purple and orange players. When the process by which the high-stake group members were selected was explained, the instructions pointed out that the reason that all the four purple players were selected was that this arrangement fulfilled the requirement of equal proportions for the high-stake group, hence giving all high-stake group members the chance to earn more money than what would otherwise have been the case.

Treatment 1c was designed to capture the fairness argument, which was done by introducing a harder math task in which the numbers that were to be added up had three digits instead of two. The instructions then pointed out that there were two math tasks, one easy and one hard, and that the four purple players would do the hard math task. They would still be compensated with 1 point (30 cents) for every correct answer. When it was explained how the selection into the high-stake group would work, it was stated that the reason that all the four pur-

ple players were put in the high-stake group, was that this was a compensation for the fact that they did a harder math task.

Treatments 1, 1b and 1c are identical in the sense that in all three cases, the four orange players, out of a total of 12, with the best performance in the math task were admitted into the high-stake group whereas the four participants who were randomly selected to be purple were automatically admitted into the high-stake group. The difference between the three treatments is that whereas there was no rationale for the preferential treatment of purples in treatment 1, treatments 1b and 1c provided, respectively, an efficiency- and a fairness justification. These two versions of treatment 1 allow us to better understand whether the effect of the unrepresentative quota on purples on cooperation in the public good game is different when a rationalization for the quota is given and, in turn, if it matters *how* the quota is rationalized.

2.2 Random Selection: Treatments 3 and 4

The unrepresentative quota, as it is designed in this experiment, captures two different aspects of how quotas are often perceived. First, that the selection process gives preferences to one group over another and second, that performance plays less, or no, role when candidates from the preferentially treated group are selected.

In order to understand the role played in generating the quota effect by these two factors, I designed treatments 3 and 4, which were identical to treatments 1 and 2 respectively, with the only difference being that the selection into the high-stake group was done randomly instead of by performance. Treatment 3 was the unrepresentative quota treatment with random selection, where the four purple players were automatically chosen for the high stake group whereas four of the twelve orange players were randomly selected. Treatment 4 was the control treatment with random selection, where four purple and four orange players were chosen out of the totally eight orange and eight purple participants.

By comparing the behavior in the public good game in the high-stake group between treatments 3 and 4, we can learn whether any differences in cooperation that are found between

treatments 1 and 2 also appear in a similar environment where selection is not done on the basis of performance.

Since treatments 3 and 4 use random selection by design and no reference is made to performance in the math task, we can also use these treatments to investigate whether there is an inherent relationship between how well a person performs on the math task and how she behaves in the public good game. This is important in order to rule out the possibility that differences in behavior between treatments 1 (and 1b and 1c) and 2 are associated with math ability and its effect on public good game behavior.

2.3 Implementation

The experiment was conducted at the Harvard Decision Science Laboratory and a total of 22 sessions were run. There were five sessions each of treatments 1 and 2, three sessions each of treatments 1b and 1c, and three sessions each of treatments 3 and 4.⁷ The sessions were conducted over the course of four days in late October and early November 2011 and five days in April 2012.⁸ All sessions had 16 participants who were recruited through the SONA system at the Harvard Decision Science Laboratory. Out of the total of 352 participants, who were only allowed to participate once, 55.4 percent were women. The median age was 23 years and the participants earned on average 23 USD (including a USD 10 show-up fee) for participating in an experimental session that lasted about 45 minutes.

The participants arrived at the lab a few minutes before the scheduled start and signed consent forms. When 16 people had arrived they were taken to the lab by the experimenter.⁹

⁷ Two additional sessions which were started had to be interrupted and canceled due to computer and network malfunctioning. Participants in these two sessions were paid for their participation and were not allowed to participate again.

⁸ In October/November 2011, treatments 1 and 2 were conducted. In April, treatments 1 and 2 were replicated and the other treatments were added. Apart from this, treatments were randomly allocated between sessions. The data on treatments 1 and 2 show no seasonal differences depending on whether data collection was done in October/November or in April. Pooled data from these two sets of sessions is therefore used in the analysis. Further details are given in Appendix C.

⁹ Over-recruitment was planned in order to reduce the probability that sessions had to be cancelled due to too few participants showing up. The 16 individuals who showed up first were allowed to participate in the

They were all seated in one room, in 16 separate cubicles. The cubicles prevented them from seeing what any other participant was doing and which color he or she had been allocated. The experiment was programmed in z-tree (Fischbacher, 2007) and instructions were given both verbally, to ensure common knowledge, and on the computer screen. Key pieces of information were also given on paper so that the participants could review these parts of the instructions at any time. Participants made all their decisions on the laboratory computers.

Instructions were given before each part of the experiment. Before the math task in part 2 and before the public good game in part 4, participants answered quizzes in order to ensure that everyone had understood the instructions correctly. The participants had to answer all questions correctly for the computer program to continue to the next screen. Those who experienced difficulties in answering any of the questions could request help by pressing a help button and they then received additional explanation in private by the experimenter. Thereafter they answered the questions in the quiz again.¹⁰

2.4. Conceptual Framework and Hypotheses Generation

I use the inequity-aversion model of Fehr and Schmidt (1999) to model the two-person public good game.

Consider two individuals, i and j , who are matched in a public good game. Each individual has an endowment y_i and decides how much of the endowment to keep and how much to contribute to a joint project (contributions by individual i are denoted c_i). The total contributions made by both individuals are multiplied by R , which is the return of the project, and the resulting amount is shared equally between i and j . The monetary payoff for individual i is hence:

$$\pi_i(c_i, c_j) = y_i - c_i + \frac{R}{2}(c_i + c_j).$$

session. Any additional participants above the first 16 received only the show-up fee of USD 10 and could not participate.

¹⁰ Note was taken of the (very few) participants who experienced substantial difficulties in understanding the instructions. Excluding these observations does not make a difference to the analysis.

If $1 \leq R \leq 2$ this is a social dilemma, i.e. setting $c_i = 0$ maximizes profit for individual i but setting $c_i = y_i$ maximizes total social payoff for the two participants. In this experimental design $R = 1.5$.

Contributions in one-shot public good games are usually well above zero (see survey of public goods studies in Ledyard, 1995). This is a puzzle if players only care about their own monetary payoffs, but not if some players care also about the equality of payoffs. Following Fehr and Schmidt I assume that the utility of individual i is

$$U_i = \pi_i - \alpha_i \max(\pi_i - \pi_j, 0) - \beta_i \max(\pi_j - \pi_i, 0).$$

This utility function captures two things; first that a deviation from equal payoffs is aversive and second that this disutility may differ depending on whether the inequality comes from individual i having a lower or a higher payoff than individual j (Fehr and Schmidt assume that $\alpha_i \leq \beta_i$, i.e. that people suffer more from inequality that is to their disadvantage).

An individual with this utility function that is deciding how much to contribute in the public good game will need to form beliefs about how much the other person will contribute. Now there is no longer a unique equilibrium with $c_i = 0$, but all contributions levels such that $0 \leq c_i \leq y_i$ can be equilibria when supported by appropriate beliefs about c_j . The extent to which participants contribute to the joint project in a public good game is viewed as a measure of cooperation, since all contributions such that $c_i > 0$ are signaling other-regarding preferences (either in the form of a higher α_i or β_i or as more positive beliefs about the contributions of the other player's contribution, c_j).

This experiment tests whether there is less cooperation in a group that is put together by an unrepresentative quota (treatment 1) than in the control treatment where all participants are treated symmetrically and selected by performance (treatment 2). That implies that I am testing the null hypothesis of no difference in public good contribution against the alternative hypothesis that there is a difference in public good contribution between these two treatments. To understand the role of justification for the quota effect, I look at contributions to the public good in the two treatments where a justification is given, treatments 1b (efficiency

justification) and 1c (fairness justification), and test the null hypothesis that they are the same as in treatment 1, where no justification was given.

Section 4 discusses different theories that could support a difference in cooperative behavior between the two treatments. As explained there, an important step to distinguish between the different mechanisms is to look at behavior broken down by the color-match of the two players, i.e. by whether the participant herself, and the person to whom she is matched, belong to the group which is disadvantaged or advantaged by the unrepresentative quota system. I hence test the null hypothesis that the difference between public good contribution in treatments 1 and 2 is the same for all color-matches against the alternative hypothesis that they are not all the same.

As discussed above, the unrepresentative quota in treatment 1 (and 1b and 1c) captures two aspects of how quotas are often perceived: preferential treatment and an asymmetric role for performance in the selection of candidates from the different groups. In order to understand the role of these features in generating the quota effect I consider contributions to the public good in treatments 3 and 4. In both of these treatments, selection is made randomly into the high-stake group. However, whereas there is an unrepresentative quota for purples in treatment 3, participants of both colors have the same chance of being selected in treatment 4. I test the null hypothesis of no difference in average contribution to the public good in these treatments.

3. Results

The experimental design, where each participant is making decisions in seven public good games, implies that there are multiple behavioral data points for each participant. It is hence important to adjust standard errors for the fact that observations for a single individual are not independent. In the analysis below, this is done by clustering standard errors on individual.¹¹ As an additional measure, in order to be as conservative as possible in testing our

¹¹ Another way of handling the data is to calculate the average for each participant, and hence only use one data point per participant. The results reported here are not sensitive to the choice of either method.

hypotheses, our tests utilize standard errors that are also clustered at the level of the experimental session in which the data were generated.¹²

3.1. Selection by performance: Treatments 1 and 2

We begin by looking at the results for treatments 1 and 2. Treatment 1 is the unrepresentative quota treatment, in which the four purple players were automatically selected for the high-stake group whereas the orange players were selected based on their performance in the math task. In the high-stake group in this treatment, average contribution in the public good game was 32.7 percent of the endowment. Treatment 2 is the control treatment in which four players of each color were selected based on performance in the math task. Under this condition, contributions in the public good game were on average 54.7 percent. These data are outlined in Figure 1.

[Figure 1 about here]

The difference in public good contribution between the two treatments is 22.0 percentage points, which in turn is highly statistically significant ($p < 0.01$ ¹³). We therefore have the following result:

Result 1: There is significantly less cooperation in the public good game when the high-stake group is put together by the unrepresentative quota than in the control treatment.

We continue by breaking down the results by whether the player is orange (and hence disadvantaged by the unrepresentative quota in treatment 1) or purple (and hence advantaged by the quota in treatment 1). Figure 2 shows these data.

[Figure 2 about here]

¹² See Fréchette (2012) for a discussion about potential session effects in laboratory experiments. The two-dimensional clustering (i.e. clustering on both individual and experimental session) that is used here is an extension of the standard cluster-robust variance estimator for one-dimensional clustering, see Cameron et al. (2011). The results reported here are not sensitive to whether clustering on session is included or excluded.

¹³ This p-value and all other p-values reported below, unless otherwise noted, come from two-sided Wald tests with standard errors clustered as described above. See Appendix E for further details.

We see that for the oranges, who were disadvantaged by the quota, we have an average contribution of 30.7 percent in treatment 1, compared to 54.2 percent in treatment 2, i.e. a quota effect of 23.5 percentage points ($p < 0.01$). For the purples, the contribution levels are 34.8 percent and 55.1 percent in treatment 1 and 2 respectively, and the quota effect is hence 20.3 percentage points ($p < 0.05$). In the control treatment, orange and purple participants behave in a way that is statistically similar and a difference-in-difference analysis reveals that the 3.2 percentage point difference in the quota effect between oranges and purples is not statistically significant. We therefore have the following result:

Result 2: In the unrepresentative quota treatment, both those disadvantaged and those advantaged by the quota cooperate less compared to the control treatment.

We continue the analysis of the results from treatments 1 and 2 by looking at whether there is a difference between how players act when they play the public good game with others who have the same color compared to those who have the other color. This can be seen in Figure 3.

[Figure 3 about here]

From the data shown in Figure 3, two conclusions can be drawn. We first note that there is less cooperation in the public good game in the unrepresentative quota treatment than in the control treatment regardless of whether participants play with someone who has the same color as themselves or with someone of the other color. The quota effect is 19.5 percentage points when participants play with others who have the same color ($p < 0.01$) and 23.8 percentage points when the game is against someone of the other color ($p < 0.01$). The difference in difference is not statistically significant. A second conclusion is that the participants display an ingroup favoritism similar to what has been found in previous research even if it is not what is driving the quota effect.¹⁴ Considering both treatments, we find that participants

¹⁴ Ingroup favoritism refers to the fact that people easily divide themselves and others into social categories and treat members of the own group (the ingroup) more favorable than people in other groups (the outgroup(s)). The conditions for this effect to arise are weak, and the effect is hence often referred to as the “minimal group paradigm” (Tajfel, 1970). See also Tajfel and Billig, 1971. Ingroup favoritism is discussed further in Section 4.1 below.

contribute significantly less (on average 3.9 percentage points, $p < 0.05$) when the person they are playing a game with does not have their color. However, even though significant, this effect is small compared to the quota effect of 22.0 percent (a Wald test where the null hypothesis is that the size of the two effects are the same size is rejected with $p < 0.01$). Our third and fourth results are hence:

Result 3: The quota effect is present and does not statistically differ in size depending on whether the matched players have the same color or not.

Result 4: Considering both the unrepresentative quota- and the control treatment, there is ingroup favoritism in the sense that participants contribute more when playing with people of the same color.

As discussed above, a key step in understanding which mechanism drives the quota effect is to examine in greater depth whether there are differences in how the people of the two different colors play with each other. This is graphed in Figure 4.

[Figure 4 about here]

Figure 4 shows that there is less cooperation in the unrepresentative quota treatment than in the control treatment, regardless of the match between the player's own color and the color of the person she plays with. The point estimate of the quota effect is largest when orange players play with purple players (24.1 percentage points) and smallest when purple players play with other purples (16.3 percentage points). However, the quota effect is statistically significant for all combinations of colors (orange to orange: $p < 0.01$, orange to purple: $p < 0.01$, purple to purple: $p < 0.1$, purple to orange: $p < 0.01$) and the differences in point estimates of the quota effect are not statistically different from one another. Our fifth result is therefore:

Result 5: The quota effect is present for all combinations of matched players' colors and not statistically different in size among the different cases.

As described in Section 2, I also asked the participants about whether they perceived the procedure through which the high-stake group was selected as fair. We will now look at the data from these questions.¹⁵ Figure 5 outlines these data for treatments 1 and 2.

[Figure 5 about here]

Figure 5 shows that 60.7 percent of participants found the process fair in the unrepresentative quota treatment, compared to 80.3 percent of participants in the control treatment. The difference of 19.6 is statistically significant ($p < 0.05$). We can therefore conclude that the unrepresentative quota was viewed as a more unfair process by the participants than the process in which everyone was selected based on performance.

3.1.1. Justification of the Unrepresentative Quota: Treatments 1b and 1c

As discussed in Section 2 some previous research suggests that the acceptance for affirmative action policies is higher when the policy is justified. I test whether justifications of the use of an unrepresentative quota have an effect on group cooperation with treatments 1b and 1c.

In treatment 1b, an efficiency argument was given as a rationalization for the use of the quota. The participants were told that the payoff rule for part 4 was such that the high-stake group would only have higher stakes than the regular-stake group if the high-stake group consisted of an equal number of orange and purple players. When the unrepresentative quota was introduced, it was done with reference to this payoff rule, pointing out that the unrepresentative quota guarantees equal representation of both colors and hence higher endowments for all members of the high-stake group. The result of this treatment in relation to treatments 1 and 2 is shown in Figure 6.

[Figure 6 about here]

¹⁵ For the fairness data we only have one observation per participant (which is that participant's assessment of the fairness of the procedure) and hence the above discussion about clustering on the level of the individual does not apply here. Reported p-values from test on fairness data come from two-sided Wald tests.

Figure 6 reveals that the contribution level in treatment 1b was 28.6 percent. This is not statistically different from the case with no justification and the difference of 26.0 percentage points between treatments 1b and the control treatment, treatment 2, is highly statistically significant ($p < 0.01$).

In treatment 1c, the unrepresentative quota was rationalized as a compensation for previous unfair treatment. The four purple participants had to do a harder math task (summing up three-digit numbers instead of two-digit numbers) without getting compensated in terms of higher payoffs per correct answer. The fact that they were all automatically selected for the high-stake group was framed as making up for this unfairness. Figure 7 below outlines the extent to which participants contributed in the public good game in treatments 1 and 1c, compared to treatment 2.

[Figure 7 about here]

As is evident from Figure 7, the fact that the quota was given with this justification did not have a positive impact on the cooperation level, compared to the situation in which the quota was left unjustified, as there is no statistical difference in cooperative behavior between treatments 1 and 1c. The difference between cooperation in treatment 1c and treatment 2 of 22.5 percentage points is however statistically significant ($p < 0.01$). The analysis of treatments 1b and 1c, in addition to the analysis of treatments 1 and 2, hence gives us the following result:

Result 6: There is less cooperation when the group is put together by an unrepresentative quota, compared to the situation where everyone is selected by performance. This difference remains even when the unrepresentative quota is justified with either an efficiency- or a fairness argument.

It is important to note that these results do not imply that there are no possible justifications that could increase cooperation in the unrepresentative quota treatment to the level that we

see in the control treatment. However, they do suggest that the negative effect of the unrepresentative quota on cooperation is quite robust.¹⁶

In Figure 8 we again look at the participants' perception of whether or not the selection process was perceived as fair and consider the two treatments in which a justification for the use of the unrepresentative quota was provided.

[Figure 8 about here]

In Figure 8 we see that the two justifications had different effects on participants. The efficiency justification given for the use of a quota in treatment 1b did not change participants' perception of equity in the process as there was no statistically significant difference between the fairness perception in treatment 1 and 1b. However, appealing to the fairness justification for the use of an unrepresentative quota in treatment 1c, the selection was perceived to be more fair ($p < 0.1$) than in treatment 1, where no reason for the quota was given.

These data add an interesting dimension to the results described above. Even though the argument used in treatment 1c worked in the sense that participants did find the process more fair, it did not have an impact on participants' behavior in the public good game, compared to the situation in which no justification was given for the unrepresentative quota.

3.2 Random Selection: Treatments 3 and 4

In order to distinguish between the roles played by the two aspects of quotas (preferential selection and an asymmetric role of performance) for generating the quota effect, I also implemented two treatments that were identical to treatments 1 and 2 with the difference that selection by performance in the math task was removed. Instead participants were randomly selected into the high-stake group. In treatment 3, which is the unrepresentative quota treatment with random selection, the four purple participants were automatically selected into the high-stake group. The four orange participants, however, were randomly selected from a

¹⁶ The reason that Section 3.1.1 does not describe the data from treatments 1b and 1c broken down by color of the player and/or the matched partner is that such an analysis does not add any new insights compared to what was discussed in the first part of section 3.1. For the interested reader the material is however available from the author.

total of twelve orange participants. In treatment 4, which is the control treatment with random selection, there were eight players of each color in each session and four of each color were randomly selected for the high-stake group. Figure 9 outlines the data from treatments 3 and 4.

[Figure 9 about here]

From Figure 9 we learn that in treatment 3, contribution to the public good was 49.3 percent of the maximum whereas it was 41.3 percent in treatment 4. This difference is not statistically significant. Furthermore, a difference-in-difference analysis reveals that the quota effect is not the same when selection is made randomly as when it is made by performance ($p < 0.01$). This gives us the following result:

Result 7: When selection by performance is removed and participants are instead selected randomly into the high-stake group, there is no negative effect of the unrepresentative quota on cooperation in the public good game.

This result highlights the importance of the selection mechanism for the quota effect to arise. When those who are disadvantaged by the quota are selected by performance in the math task there is a negative effect on cooperation from the unrepresentative quota, but this effect goes away when the selection is instead made randomly. From this we learn that a preferential selection of one color above the other is not enough to trigger lower cooperation but that the difference in selection criteria, and the role of performance, is decisive.

In Figure 10 we again ask whether participants find the process fair or unfair and consider these data for treatments 3 and 4.

[Figure 10 about here]

Figure 10 reveals that whereas 62.8 percent found the selection process fair in treatment 3, the corresponding figure for treatment 4 was 95.6 percent. This difference of 32.8 percentage points is highly statistically significant ($p < 0.01$). This is interesting as we also just noted that there was no quota effect on cooperative behavior with random selection. This further supports our conclusion that the effect that the unrepresentative quota has on cooperation is not

primarily about stated fairness perceptions. Even though participants found the unrepresentative quota in treatment 3 as unfair as in treatment 1, there was no quota effect when selection was made randomly instead of based on performance. Also, even though the participants report that they view the unrepresentative quota as more fair when the justification in treatment 1c is given, this does not change the level of cooperation compared to when no justification is given.

3.3 Math Task and Public Good Game – is there an Intrinsic Relation?

Since selection in treatments 3 and 4 is made randomly and no references to performance in the math task is made, data from these treatments make it possible to also investigate whether there is an intrinsic relationship between how a person performs in the math task and how she behaves in the public good game. Figure 11 shows a scatter plot of the number of correct answers in the math task against the percentage contribution in the public good game.

[Figure 11 about here]

Figure 11 suggests that there is no relation between the two. This is also confirmed in regression analysis (OLS) where the percentage contribution is regressed on the number of correct answers in the math task as the coefficient on number of correct answers in the math task is insignificant.¹⁷

This result tells us that the reason that we see less cooperation in treatment 1 than in treatment 2 is not a consequence of an inherent relationship between scores in the math task and behavior in the public good game.

4. Why a Quota Effect?

I have defined the quota effect as the negative impact that an unrepresentative quota has on cooperative behavior in a social dilemma, compared to a situation in which everyone in the

¹⁷ This is true also when controls are added for whether the participant was in treatment 3 or 4, the color which the participant was randomly given and whether or not she was randomly selected into the high-stake group. See Table 1 for regression results with controls.

group is treated symmetrically and selected by performance. In this section I discuss theories and previous research that provide explanations as to why we might see this effect. I start by looking at theories predicting a relation-specific quota effect and thereafter I consider research pointing in the direction of a general quota effect. Throughout, the theories are related to the results from the experiment.

4.1 A Relation-Specific Quota Effect

Several theories which can explain the existence of a quota effect also make the prediction that this should only appear in certain relations within the group, for example when those who were not advantaged by the quota interact with those who were advantaged. Here three such theories are considered.

First, there is ingroup favoritism (Tajfel, 1970; Tajfel et al., 1971), which captures people's tendency to put themselves and others into categories, and the fact that this categorization gives rise to favorable treatment of the people in the same social group as oneself compared to those in other groups (see also MacDermott, 2009; and Chen and Li, 2009). Ingroup favoritism has been found also with random groups, for example put together by a coin flip (Locksley et al., 1980). In a setting with quotas and affirmative action, the ingroup bias may be at play, since a sufficiently strong decrease in cooperation with the outgroup as a consequence of the quota could lead to a decline in overall cooperation (see also Goette et al., 2006 and Bernard et al., 2006).

However, the experimental findings in this paper are *not* predicted by ingroup favoritism. Specifically, if it were a stronger feeling of ingroup- versus outgroup-identity that was the primary difference between the control and the quota treatments, we would expect to see the quota effect more strongly when participants play with those of the other color. As we do not observe this, it can be concluded that even though there is ingroup favoritism in the sense that participants in both treatments cooperate slightly more with those of the same color, this is not what is driving the quota effect.

A second set of theories is related to entitlement and punishment. Entitlement is a key concept in the literature on procedural fairness, and it has been shown that an experimental par-

ticipant who earns an endowment in an experiment is less inclined to share those resources with others than if the endowment was randomly obtained (Hoffman et al., 1996; Cherry, 2001; Oxoby and Spraggon, 2008). Similarly, we also know that people are willing to punish those who behave in a non-approved way or who get undeserved benefits (see e.g. Fehr and Gächter, 2000; and Fehr and Fischbacher, 2004). In this case, it is not implausible that those who were not advantaged by the quota could feel an entitlement to more resources or a willingness to punish those who were advantaged by the preferential treatment. This could then give rise to the quota effect, i.e. less cooperation in the unrepresentative quota treatment compared to the control treatment.

The results that we observe in the experiment are, however, inconsistent with these mechanisms. If punishment were the main underlying factor behind the quota effect, we should not see an effect of the quota when those who were not advantaged by the quota, i.e. the orange participants, play with others of the same color. A feeling of entitlement induced by the quota treatment should on the other hand not make a difference in the case when the advantaged, i.e. purple, participants play with each other. As we observe a quota effect of similar size regardless of the participant's own color and that of the matched player, it can be concluded that the effect is not driven primarily by punishment or entitlement.

Third, it may be that (perceived) competence matters for the quota effect. Previous research has shown that affirmative action has a negative impact on how the competence of the selected person is viewed both by himself and by others (see e.g. Garcia et al., 1981; Heilman et al., 1987, 1992). There are two ways that the performance in our experiment could matter for the quota effect. On the one hand, there may be an inherent relation between how many math problems an individual can solve and how she behaves in a social dilemma. Second, there is the question of how participants view one another's competence and what beliefs they hold about whether there is a relationship between math task solving ability and cooperative behavior.¹⁸

¹⁸ See Rustichini et al., 2011, Benjamin et al., forthcoming and Mollerstrom and Seim, 2012 for examples of papers discussing the link between cognitive ability and cooperative (and other-regarding) behavior.

Returning to the results of the experiment, Section 3.3 above concluded that the quota effect is not driven by an inherent relationship between performance in the math task and behavior in the public good game. It can also be concluded that beliefs about competence are not the main driver of the quota effect, as this should not give rise to a symmetric quota effect, regardless of the color-match of the players, which is what we observe.

4.2 A General Quota Effect

There are also theories and previous research indicating that the quota effect should be general, i.e. affecting all relations within the group. We'll look at two strands of such research.

First, some previous work on affirmative action has focused on how preferential selection impacts people's general sense of interest and motivation and it has been found that when some employees or team members are selected using affirmative action, motivation, interest and social interactions can be negatively affected. Most importantly, this is true for both those who are advantaged and for those who are disadvantaged by the preferential selection (Chacko, 1982; Heilman et al., 1987, 1996. See also McFarlin and Sweeney, 1991).

Second, in the procedural justice literature, a just process is regarded as a source of social motivation, i.e. it is believed to increase cooperative behavior (Tyler and Lind, 1992; Tyler and De Cremer, 2006). This has been found in several contexts including law compliance (Tyler, 1990) and organizational citizenship behavior (Tepper and Taylor, 2003). Moreover, this effect has been shown both in the case where the outcome following the procedure was one that was favorable to the agents, and in the case where the outcome was unfavorable (see Tyler, 2003; Tyler and Huo, 2002; Zapata-Phelan, 2009; Holmvall and Bobocel, 2008).

The finding that there is a quota effect for both those who were disadvantaged and those who were advantaged by the quota, and that the effect is of similar size regardless of which of these two groups the matched player belongs to, is in line with the above findings of a general impact of procedures on behavior.¹⁹

¹⁹ Neither the procedural justice research nor the psychology literature about affirmative action has much to say regarding the exact mechanism that underlies the decrease in motivation, interest and cooperativeness that follows in the wake of unjust procedures. These results could, however, be related to the growing

5. Conclusions

Affirmative action and quotas are widely used policies that are often successful in achieving desired numerical representations. However, there are also many anecdotal examples of unwanted negative side effects in the form of uncooperative and inefficient behavior in the groups that are formed with quotas. I conduct a laboratory experiment in order to investigate whether it is indeed the case that groups put together by an unrepresentative quota cooperate less than groups where everyone is treated symmetrically and selected by performance.

I create groups by randomly allocating colors to participants. In the unrepresentative quota treatment, I create a selected group by automatically admitting all participants of one color whereas participants of the other color are chosen in competition with each other, based on their performance in a previous task. Cooperation is measured as the level of contribution in a two-person public good game and I compare the level of cooperation in the unrepresentative quota treatment to a control treatment where participants of the two colors are treated symmetrically and everyone is chosen by performance.

The results show significantly less cooperation in the unrepresentative quota treatment and I furthermore find that this effect arises both for the players who were advantaged and those who were not advantaged by the quota. The unrepresentative quota also has the same negative effect on cooperation regardless of the color of the other player. These results contradict the predictions of some prominent theories, for example that the disadvantaged will punish the advantaged.

Furthermore I find that the level of cooperation remains low in the group that is put together by the unrepresentative quota even when a justification is given for the preferential selection.

literature on how moods and emotions impact economic decision-making. It has for example been shown that people who feel anger are less other-regarding and cooperative, see e.g. van Kleef et al., 2004; Lerner and Tiedens, 2006; and Small and Lerner, 2008. One way to further deepen our understanding about why the quota effect arises in our experiment would be to utilize measures of emotions and moods. By using Likert-type scales, physiological measures or other tools, it would be possible to understand if particular emotions, such as anger or annoyance, lie behind the quota effect. See Coan et al. (2007) for an overview of the literature on emotion elicitation and assessment.

However, the negative effect of the quota on cooperation goes away when the selection criteria for the selected group is changed from performance-based to random, i.e. by removing selection by performance entirely the quota effect can be turned off. This implies that the quota effect is tied to the fact that the participants who are admitted into the selected group under the unrepresentative quota are chosen by different criteria.

There are several potential policy implications of these findings. For example, the fact that the unrepresentative quota affects all relationships in the group – not only those between the disadvantaged and the advantaged – can potentially make negative effects of affirmative action harder to detect. In organizations where affirmative action or quotas are used one often looks for differences in how people treat one another as an indication of negative effects of the policy. If no such differences are found, the conclusion may mistakenly be that all is well, even though the behavior in *all* relations in the group may be negatively affected by the policy.

Another policy lesson is that there is a difference between paying lip service to a policy and behaving in accordance with its intentions. When I introduce a fairness reason for the quota, participants move toward finding the unrepresentative quota to be more fair; in fact they find the selection process as fair as in the control treatment. However, they do not change their cooperative behavior but continue to contribute as little in the public good game as when no justification is given. This indicates that even though people may state that they find a policy to be justified or fair, its negative impact on behavior may persist.

Naturally, many laboratory findings may not translate into other settings. In group interactions outside the laboratory, for example on corporate boards, in schools and in the workplace, there are additional factors which are not captured in the laboratory. An example is the fact that relations outside the laboratory are generally of a longer duration. Also, in practice affirmative action policies are applied to groups of people who already have associations and prejudices attached to them. This may play a role in how a particular quota is perceived. However, the fact that this experiment shows an effect from preferential selection also in the abstract, short-lived and “stripped down” environment of the laboratory is an indication that this effect can arise also under minimal conditions and is thus potentially strong.

As the use of affirmative action policies and quotas spreads in many parts of the world, it is increasingly important to understand their effects on how groups function. This paper is a contribution to that research, but many questions remain unanswered. It would, for example, be interesting to conduct a version of our experiment in which measures of emotions and moods are utilized in order to investigate more specifically the effects that the unrepresentative quota have on participants.

Small alterations to the design used in this paper would also make it possible to answer questions about whether an asymmetric role of performance in itself causes less cooperation, or if it is that asymmetry together with preferential treatment that gives rise to the quota effect. Finally, in order to better assess the generality and external validity of these findings, it would be interesting to examine other group compositions than the ones used here; would the results change if the underlying group consisted of, for example, five purples and eleven oranges instead of four and twelve respectively? In order to better understand the way affirmative action and quotas impact group cooperation, I plan to address these, and other, topics in future work.

References

Ahern, Kenneth R., and Amy K. Dittmar (2012). "The Changing of the Boards: The Impact on Firm Valuation of Mandate Female Board Representation." *Quarterly Journal of Economics*, 127: 137-197.

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

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (forthcoming). "Who is 'Behavioral'? Cognitive ability and anomalous preferences." *Journal of European Economic Association*.

Bernard, Helen, Ernst Fehr, and Urs Fischbacher (2006) "Group Affiliation and Altruistic Norm Enforcement." *American Economic Review*, 96(2): 217-221.

Cameron, Colin A., Jonah B. Gelbach, and Douglas L Miller (2011). "Robust Inference with Multiway Clustering." *Journal of Business and Economic Statistics*, 29(2): 238-249.

Chacko, Thomas I (1982). "Women and Equal Employment Opportunity: Some Unintended Effects." *Journal of Applied Psychology*, 67, 119-123.

Chen, Yan, and Sherry Xin Li (2009) "Group Identity and Social Preferences." *American Economic Review*, 99(1): 431-457.

Cherry, Todd L. (2001). "Mental accounting and other-regarding behavior: Evidence from the lab." *Journal of Economic Psychology*, 22: 605-615.

Chetkovich, Carol (1997). *Real Heat: Gender and Race in the Urban Fire Service*. New Brunswick: Rutgers University Press.

Clarke, Nicola (2010). "Getting Women into Boardrooms, by Law." *New York Times*, January 27.

Coan, James A., and John J. B. Allen (2007). *Handbook of Emotion Elicitation and Assessment*. New York: Oxford University Press.

Crosby, Faye J., and Cheryl VanDeVeer (2000). *Sex, Race, and Merit: Debating Affirmative Action in Education and Employment*. Michigan: University of Michigan Press.

Dagens Naeringsliv (2010). "Går fra ASA til AS for å unngå kvinner." *Dagens Naeringsliv*, October, 13.

Daily, Catherine M, and Dan R. Dalton (2003). "Women in the Boardroom: a Business Imperative." *Journal of Business Strategy*, 24(5): 8-9.

Dreyfuss, Joel, and Charles Lawrence III (1979). *The Bakke Case: The Politics of Inequality*. New York: Harcourt Brace Jovanovich.

Fehr, Ernst, and Klaus M. Schmidt (1999). "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.

Fehr, Ernst, and Simon Gächter (2000). "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90: 980-994.

Fehr, Ernst, and Urs Fischbacher (2004). "Third-party Punishment and Social Norms." *Evolution and Human Behavior*, 25: 63-87.

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

Fréchette, Guillaume R. (2012). "Session-effect in the Laboratory." *Experimental Economics*, 15(3): 485-498.

Garcia, Luis T., Nancy Erskine, Kathy Hawn, and Susanne R. Casmay (1981). "The Effect of Affirmative Action on Attributions about Minority Group Members." *Journal of Personality*, 49(4): 427-437.

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

Goette, Lorenz, David Huffman, and Stephan Meier (2006). "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review*, 96(2): 212-216.

Heilman, Madeline E., Caryn J. Block, and Jonathan A. Lucas, (1992). "Presumed Incompetent? Stigmatization and Affirmative Action Efforts." *Journal of Applied Psychology*, 77(4): 536-544.

Heilman, Madeline E., Michael C. Simon, and David P. Repper (1987). "Intentionally Favored, Unintentionally Harmed? Impact of Sex-Based Preferential Selection on Self-Perceptions and Self-Evaluations." *Journal of Applied Psychology*, 72(1): 62-68.

Heilman, Madeline E., Winston F. McCullough, and David Gilbert (1996). "The Other Side of Affirmative Action: Reactions of Nonbeneficiaries to Sex-Based Preferential Selection." *Journal of Applied Psychology*, 81(4): 346-357.

Hoffman, Elizabeth and Matthew L. Spitzer (1985). "Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice." *Journal of Legal Studies*, 14(2): 259-297.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith (1996). "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, 86(3): 653-660.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat and Vernon L. Smith (1994). "Preferences, Property Rights and Anonymity in Bargaining Games." *Games and Economic Behavior*, 7: 346-380.

Holmvall, Camilla M., and D. Ramona Bobcel (2008). "What Fair Procedures Say about Me: Selfconstruals and Reactions to Procedural Fairness." *Organizational Behavior and Human Decision Processes*, 105(2): 147-168.

Huse, Morten, and Anne Grethe Solberg (2006). "Gender-related Boardroom Dynamics: How Scandinavian Women Make and can Make Contributions on Corporate Boards." *Women in Management Review*, 21(2): 113-130.

Huse, Morten, Sabina T. Nielsen, and Inger Marie Hagen (2009). "Women and Employee-Elected Board Members, and Their Contributions to Board Control Tasks." *Journal of Business Ethics*, 89: 581-597.

Krook, MonaLena (2009). *Quotas for Women in Politics*. New York: Oxford University Press.

Ledyard, John O. (1995) "Public Goods: A Survey of Experimental Research." In *The Handbook of Experimental Economics* (eds. John H. Kagel and Alvin E. Roth). Princeton: Princeton University Press.

Lerner, Jennifer S., and Larissa Z. Tiedens (2006). "Portrait of The Angry Decision Maker: How Appraisal Tendencies Shape Anger's Influence on Cognition." *Journal of Behavioral Decision Making*, 19: 115-137.

Lind, E. Allan, and Tom R. Tyler (1988). *The Social Psychology of Procedural Justice*. New York, NY: Plenum Press.

Locksley, Anne, Vilma Ortiz, and Christine Hepburn (1980). "Social Categorization and Discriminatory Behavior: Extinguishing the Minimal Intergroup Discrimination Effect." *Journal of Personality and Social Psychology*, 39(5): 773-783.

Lukas, J. Anthony (1986). *Common Ground: A Turbulent Decade in the Lives of Three American Families*. New York: Vintage Books.

MacDermott, Rose (2009) "Psychological Approaches to Identity: Definitions, Measurement and Experimentation." In *Measuring Identity: A Guide for Social Scientists*, (eds. Rawi Abdelal, Yoshiko M. Herrera, Alastair Iain Johnson, and Rose McDermott). Cambridge, UK: Cambridge University Press.

Matsa, David and Amalia Miller (2011). *A Female Style in Corporate Leadership? Evidence from Quotas*. Kellogg School of Management, Working paper.

McFarlin, Dean B., and Paul D. Sweeney (1992). "Distributive and Procedural Justice as Predictors of Satisfaction with Personal and Organizational Outcomes." *Academy of Management Journal*, 35(3): 628-637.

Mollerstrom, Johanna and David Seim (2012). *Does the Demand for Redistribution Rise or Fall with Cognitive Ability?*, mimeo Harvard University.

Murrell, Audrey J., Beth L Dietz-Uhler, John F. Dovidio, Samuel L. Gaertner, and Cheryl Drout (1994). "Aversive Racism and Resistance to Affirmative Action: Perceptions of Justice Are Not Necessarily Color Blind." *Basic and Applied Social Psychology*, 15(1&2): 71-86.

Niedere, Muriel, and Lise Vesterlund, (2007). "Do Women Shy Away from Competition? Do Men Compete too Much?" *Quarterly Journal of Economics*, 122(3): 1067-1101.

Niedere, Muriel, and Lise Vesterlund, (2007). "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives*, 24(2): 129-144.

Niederle, Muriel, Carmit Segal, and Lise Vesterlund (forthcoming). "How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness." *Management Science*.

Nielsen, Sabina, and Morten Huse (2010). "Women Directors' Contribution to Board Decision-Making and Strategic Involvement: The Role of Equality Perception." *European Management Review*, 7: 16-29.

Oxoby, Robert J., and John Spraggon (2008). "Mine and Yours: Property Rights in Dictator Games." *Journal of Economic Behavior and Organization*, 65: 703-713.

Pande, Rohini, and Deanna Ford (2011). *Gender Quotas and Female Leadership: A Review*. Washington DC: Worldbank.

Rustichini, Aldo, Colin G. DeYoung, Jon Anderson, and Stephen Burks (2012). *Toward the Integration of Personality Theory and Decision Theory in the Explanation of Economic and Health Behavior*. Mimeo, University of Minnesota.

Small, Deborah A., and Jennifer S. Lerner (2008). "Emotional Policy: Personal Sadness and Anger Shape Judgements about a Welfare Case." *Political Psychology*, 29(2): 149-168.

Stenseng, Sverre (2010). "Slik unngås kravet om kvinner." *Aftenposten*, October 13.

Tajfel, Henri (1970). "Experiments in Intergroup Discrimination." *Scientific American*, 223: 96-102.

Tajfel, Henrik, M.G. Billig, R.P. Bundy, and Claude Flament (1971). "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology*, 1(2): 149-178.

Tepper, Bennett J., and Edward C. Taylor (2003). "Relationships among Supervisors' and Subordinates' Procedural Justice Perceptions and Organizational Citizenship Behaviors." *Academy of Management Journal*, 46(1): 97-105.

Terjesen, Siri, Ruth Sealy, and Val Singh (2009). "Women Directors on Corporate Boards: A Review and Research Agenda." *Corporate Governance: An International Review*, 17(3): 320-337.

Tyler, Tom R. (1990). *Why People Obey the Law*. Chelsea, MI: BookCrafters.

Tyler, Tom R. (2003). "Procedural Justice, Legitimacy, and the Effective Rule of Law." *Crime and Justice*, 30: 283-357.

Tyler, Tom R. and E. Allan Lind, (1992). "A Relational Model of Authority in Groups." *Advances in Experimental Social Psychology*, 25: 115-191.

Tyler, Tom R., and David De Cremer (2006). "Cooperation in Groups," in (eds. David De Cremer, Marcel Zeelenberg, and J. Keith Murnighan) *Social Psychology and Economics*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

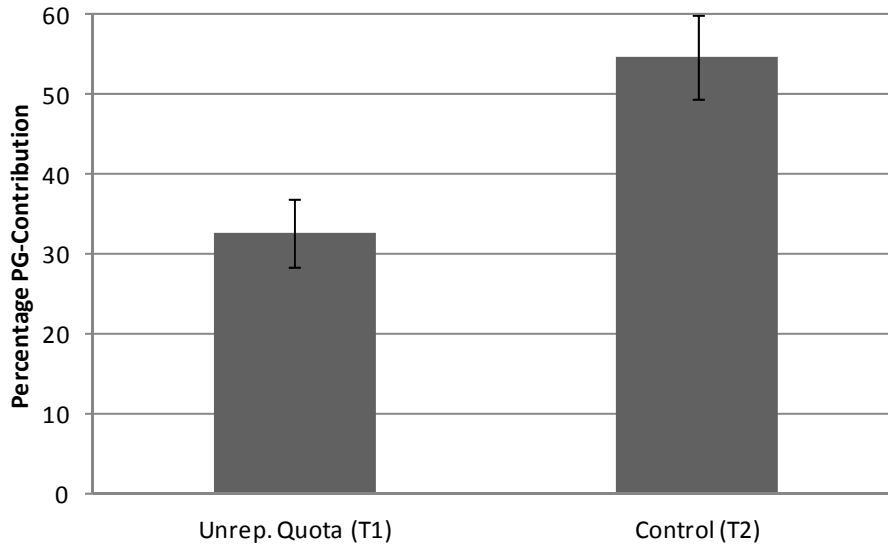
Tyler, Tom R., and Yuen J. Huo (2002). *Trust in the law*. New York: Russell Sage.

van Kleef, Gerben A., Carsten K.W. De Dreu, and Anthony S.R. Manstead (2004). "The Interpersonal Effects of Anger and Happiness in Negotiations," *Journal of Personality and Social Psychology*, 86(1): 57-76.

Zapata-Phelan, Cindy, Jason A. Colquitt, Brent A. Scott, and Beth Livingston (2009). "Procedural Justice, Interactional Justice and Task Performance: The Mediating Role of Intrinsic Motivation." *Organizational Behavior and Human Decision Processes*, 180(1): 93-105.

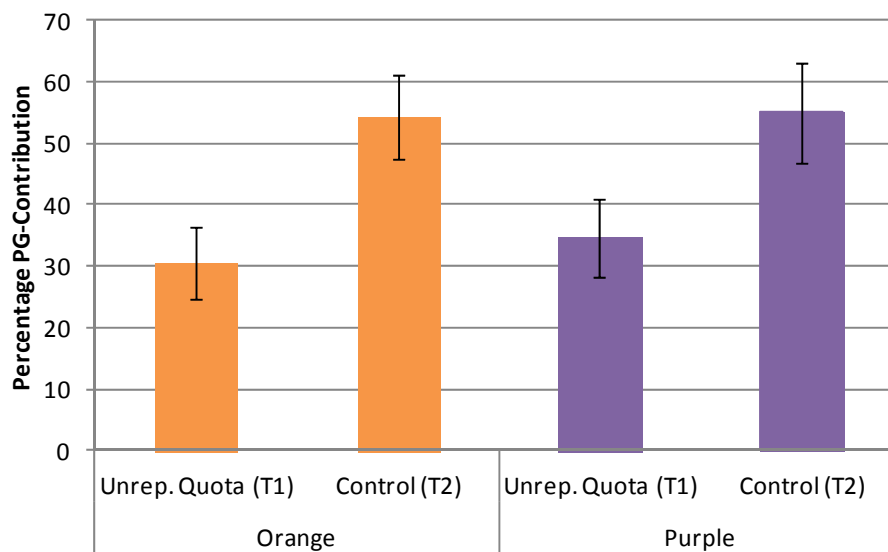
Tables and Figures

Figure 1: Contribution in PG game, High-stake group, Treatments 1 and 2.



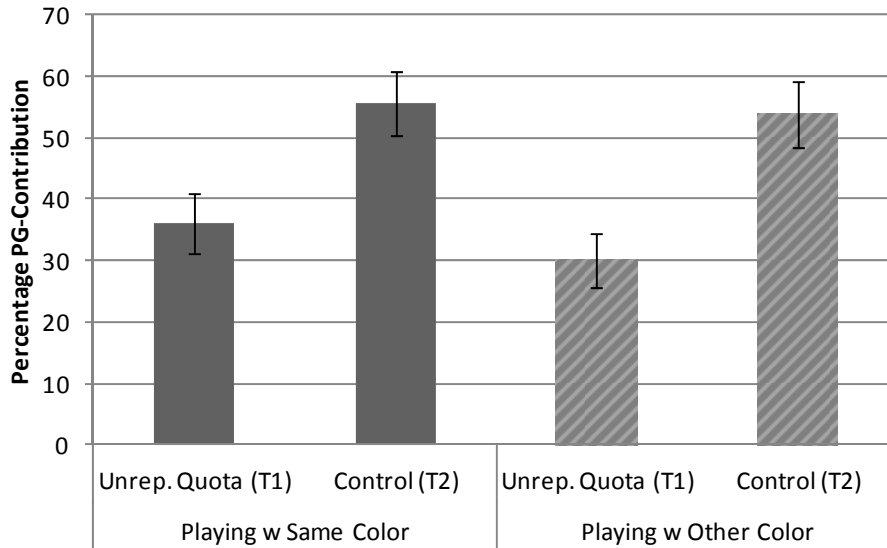
Error bars mark standard errors for average observation per participant. Number of observations: 560, number of participants: 80, number of sessions: 10.

Figure 2: Contribution PG game, High-stake group, Treatments 1 and 2, by Color.



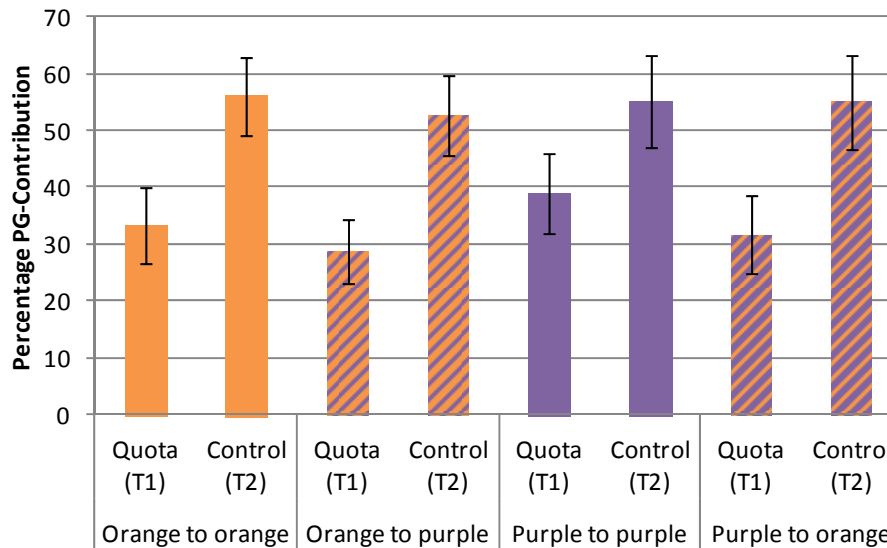
Error bars mark standard errors for average observation per participant. Number of observations: 560, number of participants: 80, number of sessions: 10.

Figure 3: Contribution in PG game, High-stake group, Treatments 1 and 2, by Color of Matched Player.



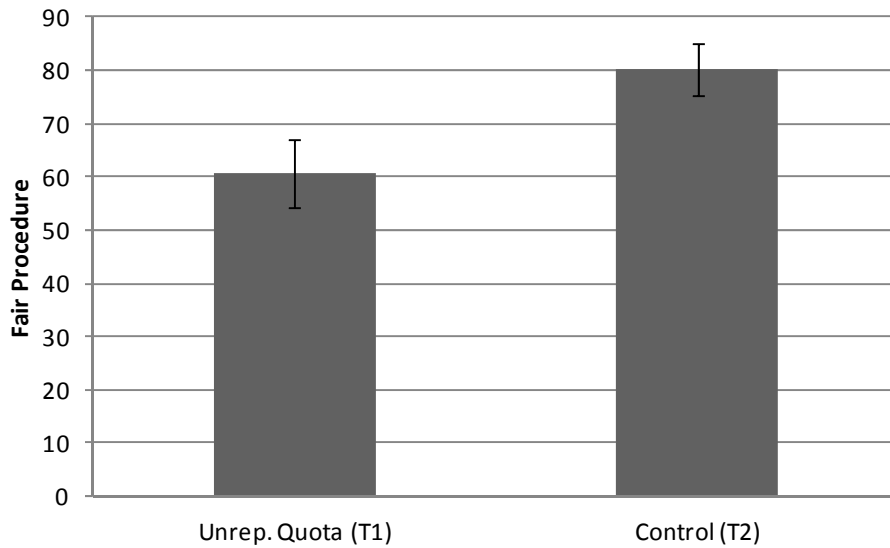
Error bars mark standard errors for average observation per participant. Number of observations: 560, number of participants: 80, number of sessions: 10.

Figure 4: Contribution in PG game, High-stake group, Treatments 1 and 2, by Color of Matched Player.



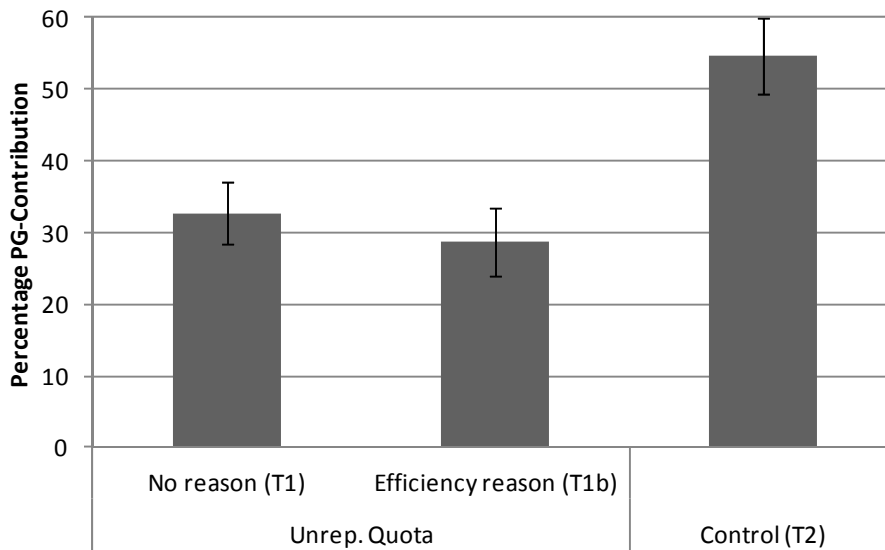
Error bars mark standard errors for average observation per participant. Number of observations: 560, number of participants: 80, number of sessions: 10.

Figure 5: Fairness Perception, Treatments 1 and 2



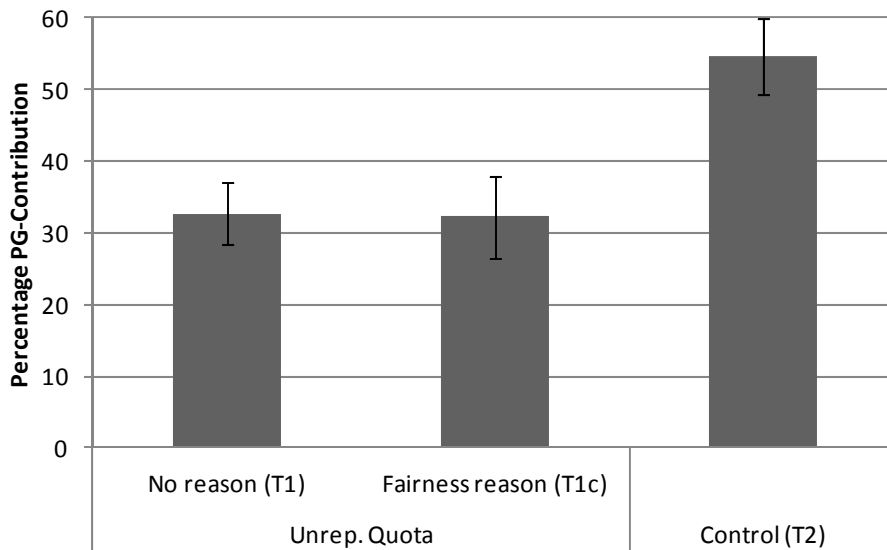
Data are from high-stake- and regular-stake group and show percentage of participants who regarded the selection process into the High-stake Group fair. Bars mark standard errors. Number of observations = number of participants: 160, number of sessions: 10.

Figure 6: Contribution in PG game, High-stake group, Treatments 1, 1b and 2.



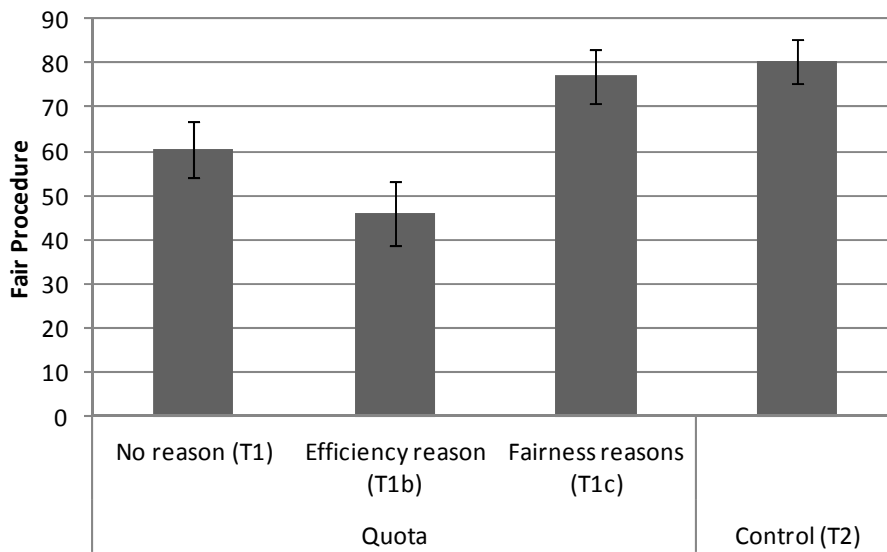
Error bars mark standard errors for average observation per participant. Number of observations: 728, number of participants: 104, number of sessions: 13.

Figure 7: Contribution in PG game, High-stake group, Treatments 1, 1c and 2.



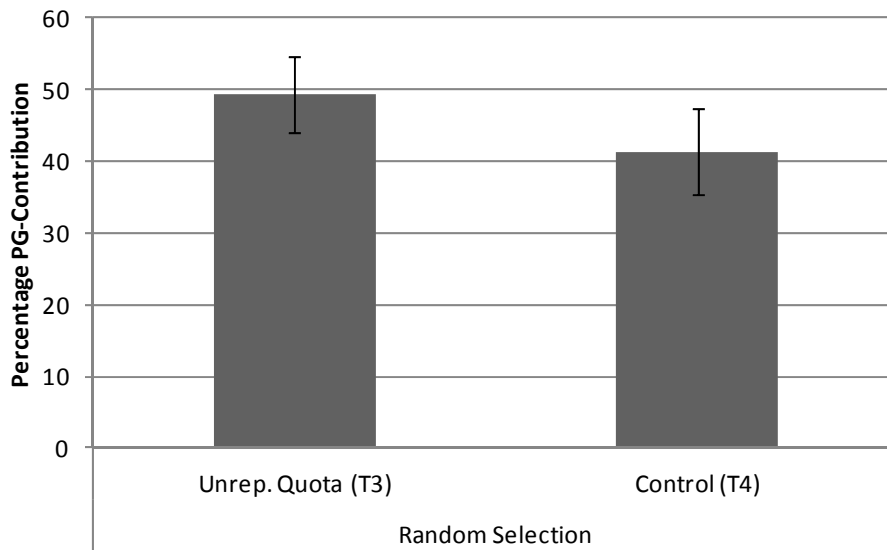
Error bars mark standard errors for average observation per participant. Number of observations: 728, number of participants: 104, number of sessions: 13.

Figure 8: The Impact of Quota Justification on Fairness Perception



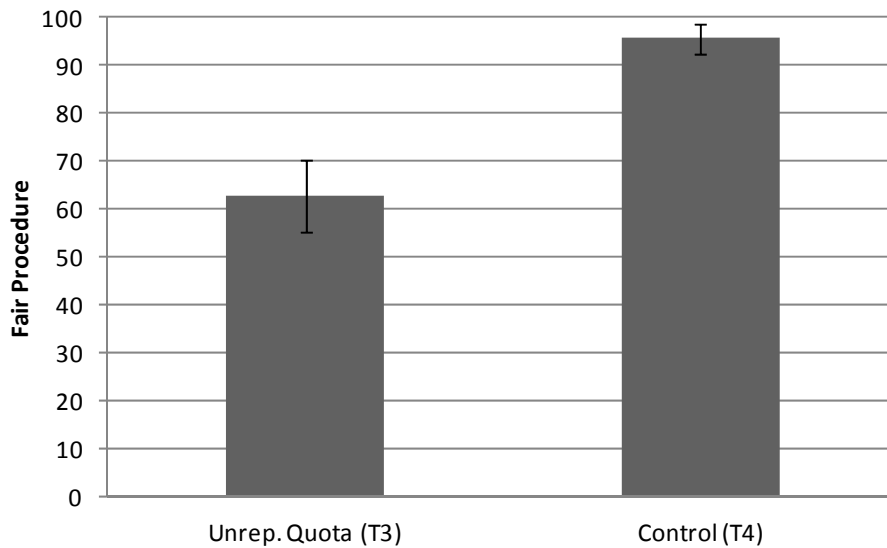
Data are from high-stake- and regular-stake group and show percentage of participants who found the selection process into the High-stake Group fair. Bars mark standard errors. Number of observations = number of participants: 256, number of sessions: 16.

Figure 9: Contribution in PG game, High-stake group, Treatments 3 and 4.



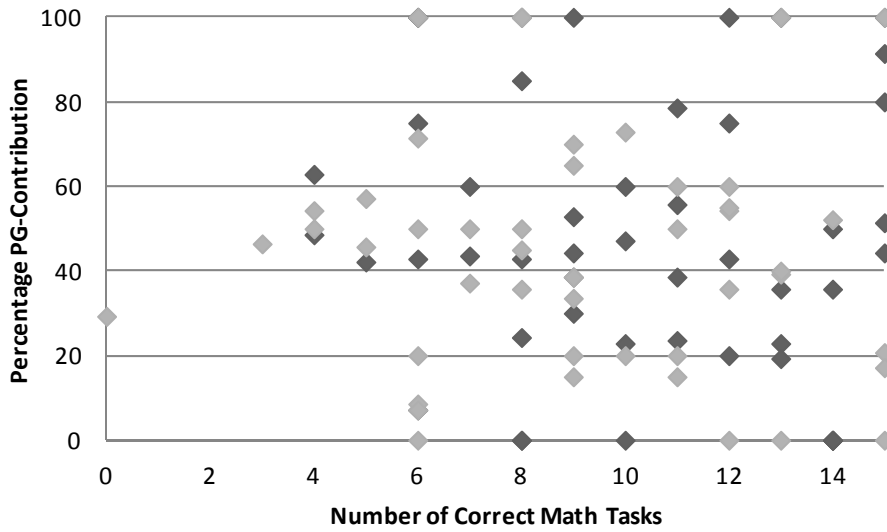
Error bars mark standard errors for average observation per participant. Number of observations: 336, number of participants: 48, number of sessions: 6.

Figure 10: Fairness Perception, Treatments 3 and 4



Data are from high-stake- and regular-stake group and show percentage of participants who found the selection process into the High-stake Group fair. Bars mark standard errors. Number of observations = number of participants: 96, number of sessions: 6.

Figure 11: Math Task Performance and Public Good Contribution, Treatments 3 and 4



Data are from high-stake- and regular-stake group. N=96. Dark gray dots denote treatment 3 and light gray dots denote treatment 4.

Table 1: Math Task Performance and Public Good Contribution, Treatments 3 and 4

	(1)	(2)
Correct	-0.007 (0.09)	-0.009 (0.10)
High Stake Group		-0.035 (0.07)
Orange		-0.05 (0.07)
Quota		0.05 (0.07)
Constant	0.528*** (0.09)	0.986* (0.55)
N	96	96

OLS. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: Percentage public good contribution. High-stake group is a dummy equal to 1 if the participant was randomly chosen as a member of the high-stake group. Orange is a dummy equal to 1 if participant is orange. Quota is a dummy equal to 1 for treatment 3.

Appendix A: Experimental Instructions

[Screen 1:]

Hi and welcome! In this study you can earn some money. The amount will depend on your decisions and the decisions of the other participants. The study has one introductory part and several parts where you can earn money. At the end of the study, your earnings (10 points correspond to \$3) will be added to the show-up fee of \$10 and you will be paid in private, in cash before you leave. We will go through the instructions now. If you have any questions after you have read and heard the instructions, please press the help button or raise your hand. Otherwise, no communication is permitted during the study. You are also not allowed to use mobile phones or other electronic devices.

[Shown in treatment 1, 1b, 1c and 3] There are 16 people, including you, participating in this study at the same time as you. You are all in this room. The 16 of you are divided into two colors: orange and purple. There are 12 orange people and 4 purple people. On the next screen you will learn which color you have.

[Shown in treatment 2 and 4] There are 16 people, including you, participating in this study at the same time as you. You are all in this room. The 16 of you are divided into two colors: orange and purple. There are 8 orange people and 8 purple people. On the next screen you will learn which color you have.

[Screen 2:]

Your color is: [ORANGE/PURPLE]. There is a total of [12/4/8/8] [ORANGE/PURPLE] people, including you. The other color is [PURPLE/ORANGE]. There are [4/12/8/8] [purple/orange] people. Please do not press the OK button until we have given you a wristband with your color.

[Screen 3:]

Use the paper and pen that are available on your desk. Write down a list of 5 associations to your color, which is [ORANGE/PURPLE]. You have a total of 2 minutes available for this task. When you have finished, press the OK button. *[Papers collected after 2 minutes and a paper with key info about next part given out.]*

[Screen 4:]

[Shown in treatment 1, 1b, 2, 3 and 4] We now move on to the first part of the study where you can earn money. In this part of the study you are asked to correctly solve as many math problems as possible. You have five minutes available. In each problem, you are asked to sum up five two-digit numbers. An example could be $32+97+13+62+20$. In this case the correct answer is 224. For each correct answer you will receive 1 point. At the end of the study you will learn how many of your answers were correct and how many points you earned. This will then be converted to dollars and paid out in cash. Please make sure to stop solving and press the OK button when

we tell you that the time limit is up. Failing to do so can negatively affect your payoffs. If you have finished all the tasks before the time limit is up, press ok and then wait until the study continues.

[Shown in treatment 1c] We now move on to the first part of the study where you can earn money. In this part of the study, the first part, you are asked to correctly solve as many math problems as possible. You have five minutes available. In each problem, you are asked to sum up five numbers. The math task can be either easy or hard. If it is easy, the numbers have two digits. An example could then be $32+97+13+62+20$. In this case the correct answer is 224. If it is hard, the numbers have three digits. An example could then be $223+761+130+409+901$. In this case the correct answer is 2424. Regardless of whether the math task is easy or hard, you receive 1 point for each correct answer. The orange players do the easy math task and the purple players do the hard math task. At the end of the study you will learn how many of your answers were correct and how many points you earned. This will then be converted to dollars and paid out in cash. Please make sure to STOP solving and press the OK button when we tell you that the time limit is up. Failing to do so can negatively affect your payoffs. If you have finished all the tasks before the time limit is up, press ok and then wait until the study continues.

[Screen 5:]

We will now check that everyone has understood the instructions for part 1 correctly! Please answer the questions on this screen. If you need help, please press the help button or raise your hand. When you have finished answering, please press "I understand". If any of your answers are incorrect, the program will tell you so and you get to answer that question again. *[Quiz is given.]*

[Screen 6:]

The math solving task will start in a few moments.

[Screen 7:]

Add up the five numbers in each row and write the sum in the box labeled "total". Please make sure to STOP solving and press the OK button when we tell you that the time limit is up. *[Math tasks on screen.]*

[Screen 8:]

It has now been determined how many of your answers were correct. At the end of the study, you will learn how many correct answers you gave and the money you earned will be given to you in cash. We now move on to part 2 of the study where you can earn more money. Please do not press OK until you have received the paper with the key information about part 2. *[Paper with key info about part 2 handed out].*

[Screen 9:]

We now move on to part 2 of the study. In this part two different groups will be formed. 8 out of the 16 people in this room will be members of a HIGH-STAKE GROUP. The other 8 will remain in the study as members of the REGULAR-STAKE GROUP. The people who are selected for the

high-stake group will have the chance to earn more money than those who are in the regular-stake group.

[Shown in treatment 1b:] The payoff rule in part 2 is as follows: If the high-stake group consists of an equal number of purple and orange participants, the payoffs in the high-stake group will be twice as large as in the regular-stake group. If the high-stake group does not consist of an equal number of purple and orange participants, the payoffs in the high-stake group will be the same as in the regular-stake group.

[Shown in treatment 1, 1b, and 1c:] The 8 high-stake group members will consist of 4 orange and 4 purple players. Since there are only 4 purple participants, the 4 purple high-stake members are simply these 4. The 4 orange high-stake members are the 4 orange participants, out of the totally 12 orange participants, who solved most math problems correctly in part 1.

[Shown in treatment 3:] The 8 high-stake group members will consist of 4 orange and 4 purple players. Since there are only 4 purple participants, the 4 purple high-stake members are simply these 4. The 4 orange high-stake members are 4 randomly chosen orange participants, out of the totally 12 orange participants.

[Shown in treatment 1b:] The reason that all the purple participants are selected as members of the high-stake group is that this group then has an equal number of orange and purple participants. Thereby the payoffs for all high-stake group members are doubled.

[Shown in treatment 1c:] The reason that all the purple participants are selected as members of the high-stake group is that they thereby are compensated for the fact that their math task was harder in Part 1.

[Shown in treatment 2:] The 8 high-stake group members will consist of 4 orange and 4 purple players. The 4 purple high-stake members are the 4 purple participants, out of the totally 8 purple participants, who solved most math problems correctly in part 1. The 4 orange high-stake members are the 4 orange participants, out of the totally 8 orange participants, who solved most math problems correctly in part 1.

[Shown in treatment 4:] The 8 high-stake group members will consist of 4 orange and 4 purple players. The 4 purple high-stake members are 4 randomly chosen purple participants, out of the totally 8 purple participants. The 4 orange high-stake members are 4 randomly chosen orange participants, out of the totally 8 orange participants.

You will shortly be informed about whether you have been selected as a member of the high-stake group or whether you remain in the study as a member of the regular-stake group. After the high-stake group has been formed, everyone will play a game. The game will be identical for everyone, but the people in the high-stake group will have the chance to earn more money. We will go through the instructions for the game in part 2 now. Please press the OK button.

[Screen 10:]

THE GAME: This game is played in pairs so you will play with one person at a time. Every person will play the game seven times with seven different people. At the end of the study, you will get paid according to the outcome of one randomly chosen game out of the seven. What you earn in that game will be converted to dollars and paid out in cash together with your other earnings. In this game both people in the pair start with an endowment of a certain number of points. Your task is to choose how much of your endowment to keep and how much to contribute to a project. The sum of the points that you, and the person you are playing with, contribute to the project will be multiplied by 1.5. The resulting number of points will then be divided equally between the two of you. Your earnings will hence be whatever you keep plus your share of the payoff from the project.

There are two differences between the high-stake group and the regular-stake group. 1. Everyone will only play with members of their own group. I.e. members of the regular-stake group will play with each other and the members of the high-stake group will play with each other. 2. The endowment is 10 points in the regular-stake group and 20 points in the high-stake group. I.e. members of the high-stake group have the chance to earn more money. Let's now look at two examples.

EXAMPLE 1: A game in the regular-stake group. Imagine that you are a player in the regular-stake group and hence you play with another member of the regular-stake group. You both have an endowment of 10 points. You contribute 4 points to the project and the person you are playing with contributes 6 points. The sum of the contributions is then $4+6=10$ points. The final payoff from the project will be $10*1.5=15$ points. This will be divided between you and the person you are paired with so that you both get $15/2=7.5$ points from the project. Since you kept 6 points out of your original 10, you will end up with $6+7.5=13.5$ points from this game. The other person kept 4 points and will get $4+7.5=11.5$ points.

EXAMPLE 2: A game in the high-stake group. Imagine that you are a player in the high-stake group and hence you play with another member of the high-stake group. You both have an endowment of 20 points. You contribute 12 points to the project and the person you are playing with contributes 8 points. The sum of the contributions is then $12+8=20$ points. The final payoff from the project will be $20*1.5=30$ points. This will be divided between you and the person you are paired with so that you both get $30/2=15$ points from the project. Since you kept 8 points out of your original 20, you will end up with $8+15=23$ points from this game. The other person kept 12 points and will get $12+15=27$ points. Please click OK.

[Screen 11:]

We will now check that everyone has understood the instructions for part 2 correctly! Please answer the questions on the screen. If you need help, please press the help button or raise your hand. When you have finished answering, please press "I understand". After everyone has finished answering the questions below, we will announce if you have been selected for the high-stake group or not. *[Quiz is given.]*

[Screen 12:]

[Shown to those selected for the high-stake group:] You have been selected as a member of the high-stake group. You will now play the game with each of the other seven members of the high-stake group.

[Shown to those not selected for the high-stake group:] You have not been selected as a member of the high-stake group. Hence you remain in the study as a member of the regular-stake group. You will now play the game with each of the other seven members of the regular-stake group.

[Screen 13:]

[Shown to those in high-stake group:] You are a member of the high-stake group and you will play the game with each of the other 7 members of the high-stake group. All 7 games are carried out at the same time. The game is played anonymously and the only thing you know about the person you play a certain game with is that person's color. In each game you have 20 points. You have to decide how many points to keep and how many to contribute to the project. When you have made your choice in all games, please press OK.

[Shown to those in regular-stake group:] You are a member of the regular-stake group and you will play the game with each of the other 7 members of the regular-stake group. All 7 games are carried out at the same time. The game is played anonymously and the only thing you know about the person you play a certain game with is that person's color. In each game you have 10 points. You have to decide how many points to keep and how many to contribute to the project. When you have made your choice in all games, please press OK.

[Screen 14:]

All 7 games have now been conducted and one game has been chosen randomly for payment. You will learn the outcome at the end of the study, just before your earnings are paid out in cash. We now move on to part 3, which is the last part of the study. Please press OK to proceed to part 3.

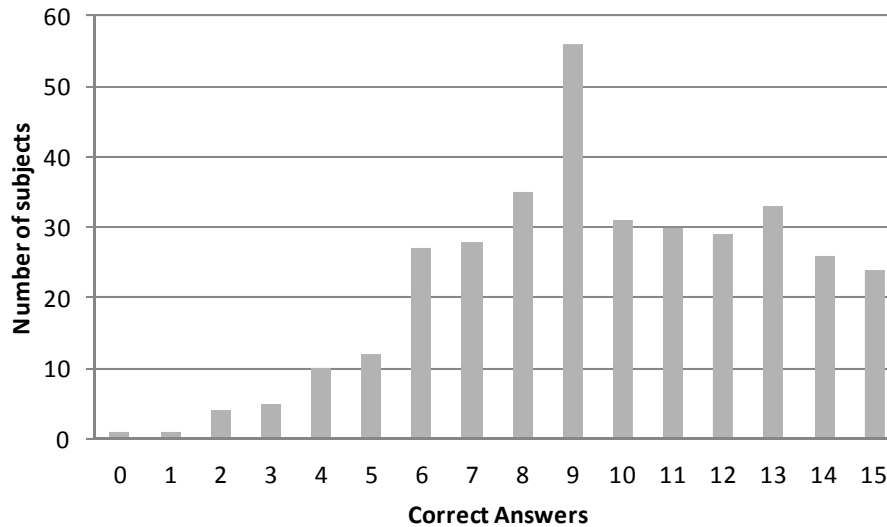
[Screen 15 to end:]

While we prepare your payments, please answer a few questions.

[Questions about fairness perception, gender, age, ethnicity, etc given]

Appendix B: Math Task Performance

Figure A-1: Distribution of Math Task Performance



Number of participants with specific number of correct answers to math task. Number of participants: 352.

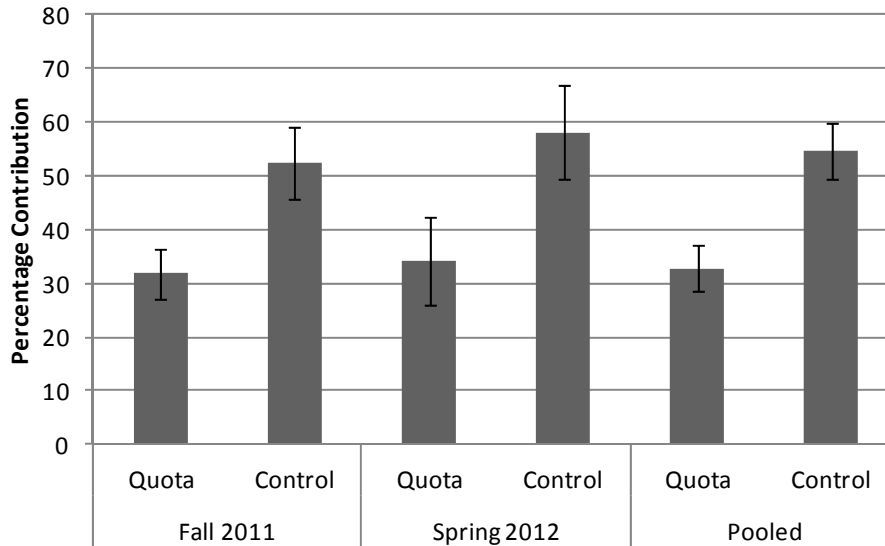
Table A-1: Correlates of Math Task Performance

	(1)	(2)	(3)	(4)
Female	-0.70** (0.34)			-0.76** (0.34)
Age		-0.10** (0.04)		-0.10** (0.04)
Non-white			0.32 (0.34)	0.37 (0.34)
Three-digits	-2.55*** (0.93)	-2.42** (0.93)	-2.48*** (0.94)	-2.42*** (0.93)
Constant	10.15*** (0.26)	12.08*** (0.98)	9.26*** (0.23)	12.26*** (1.02)
N	352	352	352	352

*OLS. Robust standard errors in parentheses. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: Number of correct math tasks. Female: dummy equal to 1 if female. Age: age in years. Non-white: dummy equal to 1 if participant has ethnicity other than white. Three-digits: dummy equal to 1 if participant solved math task with three digits (true for $N=12$ in treatment 1b).*

Appendix C: Sessions Fall 2011 and Spring 2012

Figure A-2: Results of treatments 1 and 2: fall 2011, spring 2012 and pooled



Error bars mark standard errors for average observation per participant. Fall 2011: Number of observations: 336, number of participants: 48, number of sessions: 6. Spring 2012: Number of observations: 224, number of participants: 32, number of sessions: 4.

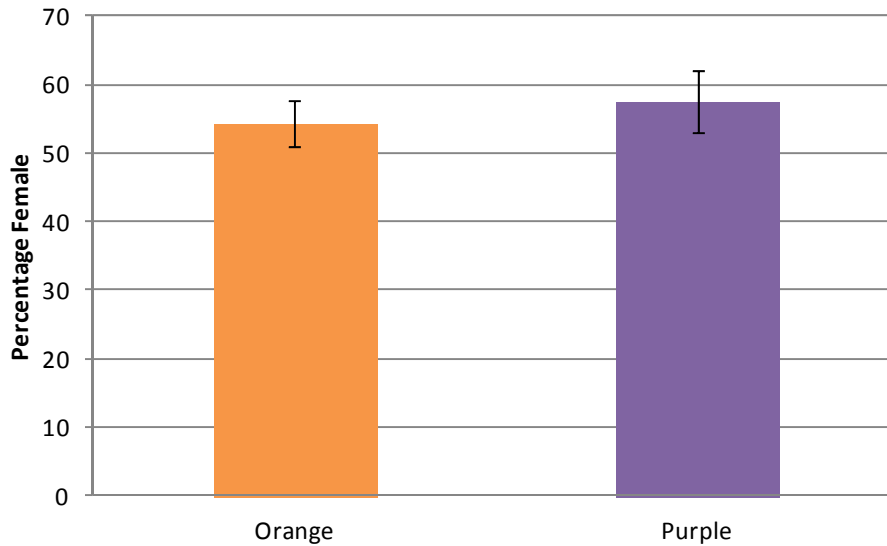
Table A-2: Regression results of treatments 1 and 2: fall 2011, spring 2012 and pooled

	(1)	(2)	(3)	(4)
Quota	-0.21** (0.08)	-0.24* (0.12)	-0.22*** (0.07)	-0.21** (0.08)
Spring				0.06 (0.11)
Quota*Spring				-0.03 (0.14)
Constant	0.52*** (0.06)	0.58*** (0.09)	0.55*** (0.05)	0.52*** (0.07)
N (cluster)	336 (48)	224 (32)	560 (80)	560 (80)

OLS. Robust standard errors in parentheses. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. (1): Only data from fall 2011. (2): Only data from spring 2012. (3) and (4): Pooled data.

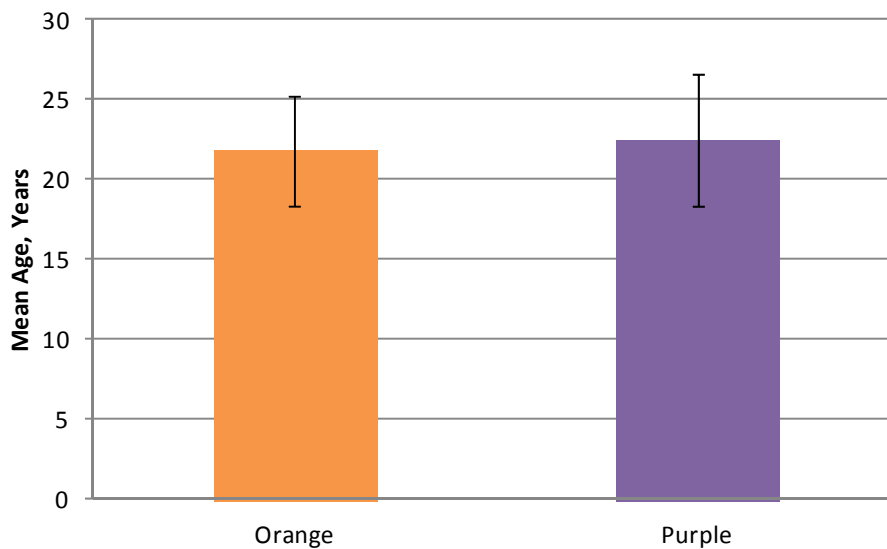
Appendix D: Orange and Purples

Figure A-3: Percentage Female, Orange and Purple



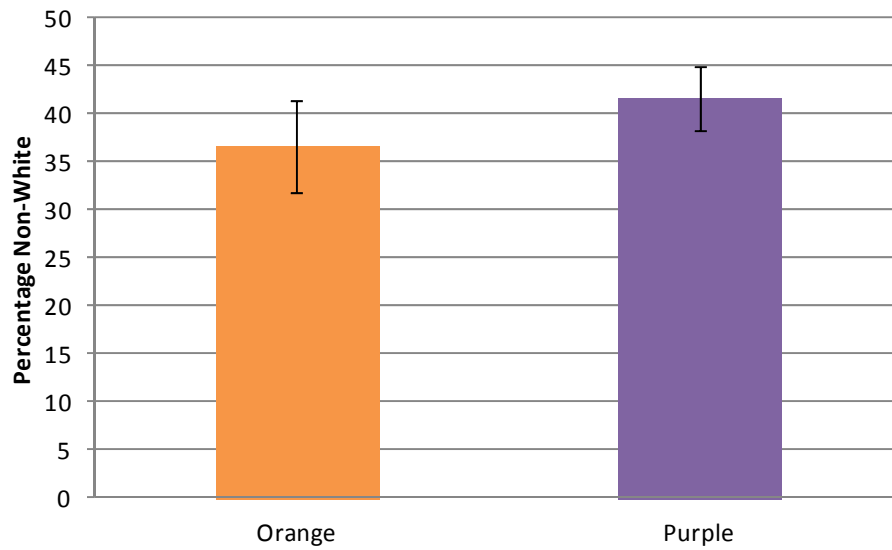
Error bars mark standard errors. No difference between colors ($p > 0.1$ in two-sample t -test with equal variances). Number of participants: 352.

Figure A-4: Mean Age, Orange and Purple



Error bars mark standard errors. No difference between colors ($p > 0.1$ in two-sample t -test with equal variances). Number of participants: 350.

Figure A-5: Percentage Non-White, Orange and Purple



Error bars mark standard errors. No difference between colors ($p > 0.1$ in two-sample t -test with equal variances). Number of participants: 313.

Appendix E: Regression Tables

Table A-3: Regressions Treatment 1 and 2, part 1

Dependent variable: Percentage contribution in PG-game									
	All HS (1)	Orange HS (2)	Purple HS (3)	To same HS (4)	To diff HS (5)	O to O HS (6)	O to P HS (7)	P to P HS (8)	P to O HS (9)
Quota	-0.2194643*** (0.0363049)	-0.2353571*** (0.0394357)	-0.2035714** (0.0799323)	-0.1954167*** (0.0475174)	-0.2375*** (0.0313234)	-0.2283333*** (0.0375582)	-0.240625*** (0.0466552)	-0.1625* (0.0869258)	-0.234375*** (0.0828503)
Multi-cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs (N)	560	280	280	240	320	120	160	120	160
Cluster (ind)	80	40	40	80	80	40	40	40	40
Cluster (sessions)	10	10	10	10	10	10	10	10	10

*O*L*S*. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error in parentheses. Cluster as indicated in table.

Table A-4: Regressions Treatment 1 and 2, part 1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: Percentage contribution in PG-game								
Quota	-0.2194643*** (0.0363049)	-0.2353571*** (0.0394711)	-0.2194643*** (0.0363375)	-0.2375*** (0.0313586)	-0.2170455*** (0.0423961)	-0.211*** (0.0571418)	-0.235*** (0.0292869)	-0.2135*** (0.0309376)
Purple		0.0089286 (0.0887349)						
Quota*Purple		0.0317857 (0.1030826)						
Samecolor			0.0386458** (0.0161641)	0.0176042 (0.0144428)				
Quota*Samecolor			0.0420833 (0.0291402)					
Orange_to_orange					0.0191667 (0.0373125)			
Quota*Orange_to_orange					-0.0112879 (0.04818)			
Orange_to_purple						-0.02675 (0.0792236)		
Quota*Orange_to_purple						-0.029625 (0.0893928)		
Purple_to_purple							0.0064394 (0.0485828)	
Quota*Purple_to_purple							0.0725 (0.0757208)	
Purple_to_orange								0.005625 (0.0701816)
Quota*Purple_to_orange								-0.020875 (0.0804317)
Multi-cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs (N)	560	560	560	560	560	560	560	560
Cluster (ID)	80	80	80	80	80	80	80	80
Cluster (session)	10	10	10	10	10	10	10	10

*OLS. Levels of significance: * p<0.1, ** p<0.05, *** p<0.01. Robust standard error in parentheses. Cluster as indicated in table.*

Table A-5: Regressions Treatment 1, 1b, 1c and 2

Dependent variable: Percentage contribution in PG-game

	<i>T1 and T1b</i>	<i>T1b and T2</i>	<i>T1 and T1c</i>	<i>T1c and T2</i>	<i>T1, T1b, T1c, T2</i>
	(3)	(4)	(1)	(2)	(5)
Quota		-0.2602976*** (0.0500737)		-0.2245833*** (0.0763087)	-0.2194643*** (0.0355993)
Effttreat	-0.0408333 (0.0458854)				
Fairtreat			-0.005119 (0.0736282)		
Effttreat*Quota					-0.0408333 (0.0443543)
Fairtreat*Quota					-0.005119 (0.0711714)
Multi-cluster	Yes	Yes	Yes	Yes	Yes
Nr of obs	448	448	448	448	896
Nr of cluster (id)	64	64	64	64	128
Nr of cluster (session)	8	8	8	8	16

OLS. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error in parentheses. Clustered as indicated in table.

Table A-6: Regressions Treatment 1, 2, 3 and 4

Dependent variable: Percentage contribution in PG-game

	<i>T3 and T4</i>	<i>T1, T2, T3, T4</i>
	(1)	(3)
Quota	0.0797619 (0.0929517)	-0.2194643*** (0.0355993)
Random		-0.1332143 (0.0869024)
Random*Quota		0.2992262*** (0.0946053)
Multi-cluster	Yes	Yes
Nr of obs	336	896
Nr of cluster (id)	48	128
Nr of cluster (session)	6	16

OLS. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error in parentheses. Clustered as indicated in table.

Table A-7: Regressions on Fairness Perceptions, all treatments

Dependent variable: Fairness perception

	<i>T1 and T2</i>	<i>T1 and T1b</i>	<i>T1 and T1c</i>	<i>T3 and T4</i>
	(1)	(3)	(2)	(4)
Quota	-0.1964729** (0.079424)			-0.3276486*** (0.0795196)
Fairtreat			0.164276* (0.089525)	
Efftreat		-0.148224 (0.0959723)		
N	127	109	109	88

*OLS. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error in parentheses.*