Replication of

Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory

by Balafoutas, L. / Sutter, M. (2012)
in: Science, 335, pp. 579–582

Replication Authors:

Felix Holzmeister, Jürgen Huber, Michael Kirchler, and Julia Rose

In a set of controlled laboratory experiments, Balafoutas and Sutter (2012) study the effects of different affirmative action policy interventions to encourage women's choice to enter competitions. Four different interventions are investigated: quotas, where one of two winners of a competition must be female; two variants of preferential treatment, where a fixed increment is added to women's performance; and repetition of the competition, where a second competition takes place if no woman is among the winners. Compared with no intervention, all interventions encourage women to enter competitions more often and performance is at least as good both during and after the competition.

Hypothesis to replicate and bet on:

With preferential treatment of women — i.e., each woman's performance is automatically increased by one unit in the competition — more women will choose to compete (a comparison of the fraction of women who chose the tournament scheme rather than the piece rate scheme in the 'preferential treatment one (PT1)' versus the 'control treatment (CTR)'; $\chi^2(1) = 5.62$, $p = 0.018$, p. 580).

(This hypothesis was picked by lottery instead of comparing PT2 to CTR; $\chi^2(1) = 10.89$, $p = 0.001$, p. 580).

Power Analysis and Criteria for Replication: First Data Collection

The original sample size is 144 participants and the standardized effect size measured as the correlation coefficient (r) is 0.197. To have 90% power to detect 75% of the original effect size, a sample size of 485 is required. The criteria for replication are an effect in the same direction as the original study and a p-value < 0.05 (in a two-sided test).

Power Analysis and Criteria for Replication: Second Data Collection

If the original result is not replicated in the first data collection, a second data collection is carried out. To have 90% power to detect 50% of the original effect size in the pooled sample (first and second data collection), a sample size of 1099 is required, i.e., a sample size of 614 in the second data collection is required. The criteria for replication are an

effect in the same direction as in the original study and a *p*-value < 0.05 (in a two-sided test) in the pooled data.

Sample

The sample in the first data collection consists of 485 students from the University of Innsbruck. If the original result is not replicated in the first data collection (two-sided *p*-value < 0.05 in the same direction as the original study), a second data collection consisting of 614 additional students from the University of Innsbruck will be carried out such that the pooled sample size is 1099. Subjects who participated in the experimental sessions of the original studies are excluded from recruiting.

Materials

We use the software of the original experiment programmed in z-Tree (Fischbacher, 2007) along with the original German instructions which have been made available by the authors.

Procedure

We follow the procedure of the original study, with only slight but unavoidable deviations as out-lined below. The following summary of the experimental procedure is therefore based on the explanations of the experimental procedure in the article (pp. 579–80) and the section "Notes on the experimental procedure" (p. 3–4) of the Supplementary Information.

Subjects are randomly assigned into groups of three men and three women. All groups go through several stages. The experimental task in each of the stages 1 to 4 is to add as many sets of five two-digit numbers as possible within 3 minutes. Ties between participants are broken randomly in stag-es 2, 3, and 4. The task in stage 5 is a simple coordination game. At the beginning of the experiment, subjects are informed about the number of stages but not about what the tasks in each of the stages will be. The instructions for each of the task are provided just before every new stage.

In stage 1 (piece rate), each subject receives €0.50 for each correct calculation. In stage 2 (tournament), group members compete against each other. The two members who solve the most calculations correctly are paid €1.50 per calculation. The other four group members receive nothing. Subjects do not receive any feedback on the outcome of the competition in stage 2 until the end of the experiment to avoid that subjects condition their choices on previous outcomes of a competition. In stage 3 (choice), subjects choose whether they want to solve the calculations under a piece rate scheme or a tournament scheme. If the tournament is chosen, a subject's performance in stage 3 is compared with the other group members' performance in stage 2. In this stage, the competition rule across the two treatments PT1 and CTR are varied to examine the effects of the policy intervention: In the control treatment (CTR), the winners are the two group members with the largest numbers of correct calculations, regardless of gender; in the preferential treatment (PT1), each woman's performance is automatically increased by one unit (i.e., one correct calculation). Subjects do not receive any feedback on the outcome of the optional competition in stage 3 until the end of the experiment in order to avoid that subjects condition their choices on previous outcomes. At the end of stage 3, beliefs of all subjects regarding their relative performance and their ranks in stages 1 and 2 are elicited. For each stage, subjects have to indicate their expected rank within the group of six members and within their own gender only. Correct guesses are rewarded with €1.00

each, and the feedback is given also only at the end of the experiment. In stage 4 (tournament with policy intervention), all subjects compete against each other's performance in stage 4 and the competition rules in stage 3, outlined above. At the end of stage 4, subjects are informed about the outcome of the competition, but not about the performance of competitors, before moving on to the coordination task in stage 5. Each winner in stage 4 receives an additional €5.00 as an initial endowment in stage 5, and each loser receives only €2.00. The reason for this unequal payment is to introduce a clear distinction between winners and losers before starting with the post-competition stage. In stage 5, each group member plays a two-person coordination game with each of the other five group members. Before playing the coordination game, a subject is informed about the gender of the other player and whether this player has won or lost in the tournament of stage 4. With this information, each subject has to choose five times a number for the interaction with each of the other group members. All decisions are made simultaneously. In stage 6, subjects have to add up two-digit numbers again. The payment scheme is such that each correct calculation is worth €0.50 for the group in total and then split equally among all group members. Total group performance in this team task can be used as an additional indicator of the impact of the intervention on lab task outcomes after the competition has been concluded. In half of the sessions in each treatment, subjects play stage 6 before stage 5 in order to control for possible order effects of the team task and the coordination game.

In order to avoid wealth effects, one stage among stages 1 to 4, and one stage among stages 5 and 6 are randomly selected for payment at the end of the experiment. The flat payment of €5.00 (€2.00) for winners (losers)

in stage 4 is paid for sure. Each subject also receives a show-up fee of €3.00.

Analysis

The analysis will be performed exactly as in the original study. That is, a χ^2-test on the difference between the frequency of women choosing the tournament scheme rather than the piece rate scheme in the 'preferential treatment one (PT1)' and the 'control treatment (CTR)' will be conducted.

In the original study, the frequency of women opting for competition was 30.6% in the 'control treatment (CTR)' compared to 58.3% in the 'preferential treatment one (PT1)'. A χ^2-test revealed that the difference between the frequency of competing women in the two treatments is statistically significant ($\chi^2(1) = 5.62$, $p = 0.018$). The same test will be used in the replication study.

The results will first be estimated based on the first data collection. If the original result is replicated in the first data collection (a two-sided p-value < 0.05 in the same direction as in the original study), the second data collection will not be carried out. If the original result is not replicated in the first data collection a second data collection will be carried out. The above statistical test will then be estimated for the pooled sample of the first and second data collection to test if the original result replicated (a two-sided p-value < 0.05 in the same direction as in the original study).

Differences from Original Study

The replication procedure is identical to that of the original study, with some unavoidable deviations. The replication will be performed at the University of Innsbruck in Innsbruck in the between September 2016 and September 2017, while the original data was gathered at the University of Innsbruck in

2009. The experiment will be in German as in the original study.

The original study investigates three more interventions in separate treatments: quotas (where one of two winners of a competition must be female), a preferential treatment (where 2 units are added to women's performance), and repetition of competition (if no women is among the winners in the first round). For the replication, the focus is only on the preferential treatment with 1 unit added to women's performance (PT1). This hypothesis was picked by lottery instead of comparing PT2 and the control treatment.

Replication Results for the First Data Collection (90% power to detect 75% of the original effect size)

[To be added when replication experiments have been completed.]

Replication Results for the First and Second Data Collection Pooled (90% power to detect 50% of the original effect size)

[To be added when replication experiments have been completed.]

Unplanned Protocol Deviations

[To be added when replication experiments have been completed.]

Discussion

[To be added when replication experiments have been completed.]

References

Balafoutas, L. / Sutter, M. (2012): *Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory*, Science, 335, pp. 579–582.

Fischbacher, U. (2007): *z-Tree: Zurich Toolbox for Ready-Made Economic Experiments*, Experimental Economics, 10(2), pp. 171–178.