Affirmative Action and Stereotype Threat

Anat Bracha
The Federal Reserve
Bank of Boston

Alma Cohen
Harvard Law School
Tel Aviv University
and the NBER

Lynn Conell-Price
Carnegie Mellon University

Abstract

This paper contributes to understanding the potential effects of affirmative action (AA) policies on behavior. In particular, we investigate whether AA policies can trigger a "stereotype threat effect" – that is, cue a negative stereotype that leads individuals to conform to the stereotype and adversely affects their performance. Stereotype threat has been shown to be potentially significant for individuals who identify with the domain of the stereotype and who engage in complex stereotype-relevant tasks. We investigate this question experimentally in the context of gender-based AA, and find an overall negative effect of AA on the performance of high-ability women performing complex math tasks. This is consistent with the hypothesis that AA policies may trigger a stereotype threat effect and thereby have an unintended negative effect on the performance of intended beneficiaries.

JEL Classifications: C91, I28, J16, J78, K19, K31
Keywords: Affirmative action, stereotype threat, gender differences, GRE performance.

We would like to thank Uri Gneezy, Muriel Niederle, Tali Regev, Analia Schlosser, Lise Vesterlund and participants of the North American ESA Meetings in Santa Cruz and the 2014 SITE experimental economic workshop for helpful comments. The project was made possible by the generous financial support of the Foerder Institute for Economic Research and the Harvard Law School. This research has been conducted with IRB approval. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Boston or the Federal Reserve System.

1. Introduction

Affirmative action (AA) has long been a hotly debated subject. This debate should be informed by an improved understanding of the consequences of AA, particularly its effect on the performance of targeted individuals. This paper provides experimental evidence that contributes to obtaining such an understanding. In particular, we provide the first evidence that AA may trigger a stereotype threat effect using a lab experiment of gender-based AA.

The topic of gender-based AA is timely; in spite the fact that women today represent half of all college-educated individuals, they are still substantially underrepresented in many selective, high-level professional positions and occupations. For example, women are starkly under-represented among corporate directors and top executives; women constitute less than 20% of corporate directors, and their fraction of top executive positions in large public companies is below 5% in both the US and EU countries (Catalyst Census: Fortune 500 Women Board Directors 2013 and European Commission 2012). Similarly, the percentage of females among partners in US law firms is much smaller than their percentage among law firm associates (National Association for Law Placement 2013), and women in OECD countries are under-represented among those with jobs in science, technology, engineering, and math (STEM) (see Meeting of the OECD Council at Ministerial level 2011 and, for the US, see U.S. Department of Commerce, Economics and Statistics Administration 2011).

This underrepresentation of women in some professional groups and positions has led to strong interest in measures aimed at increasing female representation in those capacities, including the adoption or consideration of AA policies. For example, in 2003 Norway became the first country to pass gender representation law for corporate boards, which required about 500 firms to raise the proportion of women on their boards to 40%. Belgium, France, Iceland, Italy, Malaysia, the Netherlands, and Spain, followed Norway in adopting such measures, and more recently, the European Commission has proposed legislation to ensure that, by 2020, 40% of non-executive directors will be women (European Commission 2012).

While AA policies can lead to the increased selection of women, holding performance level constant, it is important to recognize that these policies can also influence the performance on which selection is based. As a substantial literature has already shown, AA policies can counter the tendency of women to avoid competing vigorously in mixed-sex situations (see, e.g., Gneezy et al. 2003, Niederle and Vesterlund 2007, Niederle and Vesterlund 2010, Sutter and Rützler 2010, Niederle et al. 2013, and a review by Croson and Gneezy 2009). This tendency (to avoid or not compete vigorously in competition) might lead talented women either to fail to win participation in certain professional groups and positions or to avoid the selection process altogether, taking alternative career routes instead. The countering of such tendencies by AA policies, which we label "the competition effect," has been identified experimentally by several important studies (Schotter and Weigelt 1992, Calsamiglia et al. 2013, Niederle et al. 2013, and Balafoutas and Sutter 2012). These studies, using a range of simple and mostly neutral (non-stereotypical) tasks, find that AA policies have the potential to counter such tendencies and improve the performance of the intended beneficiaries.

We argue, however, that in real-life settings, in addition to its competition effect, AA policies may affect women's performance in yet another way – by acting as a prime that triggers stereotype threat. That is, AA may remind individuals that they belong to a group that stereotypically underperforms in the judged task and that, in turn, may lead them to conform to the negative stereotype. Stereotype threat, which in this paper is introduced to the analysis of AA, is the subject of a significant body of study showing that reminders of belonging to a group that stereotypically performs worse on a task adversely affects individuals' performance on such a task, especially if it is difficult and complex. For example, studies document that reminding African-American students of their race leads them to perform significantly worse on verbal Graduate Record Exams (GREs) (Steele and Aronson 1995) and that reminding women of their gender impairs their performance in math (Shih et al. 1999, Spencer et al. 1999).[1]

---

[1] At the same time, Shih et al. (1999) show that reminding Asian women of their race leads them to perform significantly better. This study therefore demonstrates the more general concept of stereotype susceptibility whereby positively stereotyped groups can also be primed and will subsequently perform better. We also

The stereotype threat effect operates through priming—that is, by reminding individuals of negative stereotypes relevant to them or by providing information about relevant differences in achievement across groups (see, e.g., Stricker 1998 and Spencer et al. 1999). AA, by definition, singles out disadvantaged or underperforming groups and therefore may unintentionally remind its beneficiaries of relevant negative stereotypes. We therefore hypothesize that the introduction of an AA plan can produce a stereotype threat, and we test this hypothesis below.

Stereotype threat effects typically occur when a task is complex and a negative stereotype is relevant to the task. Studies that have focused on testing for a stereotype threat have therefore used tasks such as the Scholastic Achievement Test (SAT) and the GRE (Steele and Aronson 1995 and Spencer et al. 1999; see Wei 2009 for a summary of lab results). Experimental studies of AA programs, however, have thus far focused on the competition effect and not examined contexts with real, complex, and stereotype-relevant tasks. For example, Schotter and Weigelt (1992) use an abstract task where (cost) disadvantage is randomly assigned; Calsamiglia et al. (2013) use performance on Sudoku puzzles across two primary schools in Spain, where one school had Sudoku training. The AA in their study targeted students in the "disadvantaged" school without prior training; yet, the training was coincidental so there is no established stereotype with Sudoku performance. Similarly, in the seminal paper of Niederle et al. (2013) they use a simple math task of summing five 2-digit numbers in a given amount of time. Although the task is in math, and people tend to associate any math tasks with negative gender stereotype, it has been shown, in fact, that for gender gap in math performance to emerge- (see Hyde et al. 1990) and for stereotype threat effect to be present - the tasks used need to be more complex. It could well be. that if Niederle et al. (2013) would have used a more complex math tasks, such as SAT and GRE (tests that are used in admission to schools), AA would have triggered a stereotype threat effect. We believe that it is important to explore which tasks have the potential to trigger stereotype threat and which one do not have the potential of triggering stereotype threat.

---

note that, although most studies in the stereotype threat literature find a stereotype threat effect, others find opposite effects (Wei 2009, 2012) or no effects (Stricker 1998, Fryer et al. 2008).

Hence, while the tasks used in the above studies enabled researchers to test the competition effect, which was the focus of their studies, the tasks did not enable them to test for a stereotype threat effect. We therefore use tasks that facilitate such testing—in particular, the solving of complex GRE quantitative questions, an endeavor in which the average performance of women is both stereotypically and actually worse than that of men. To the best of our knowledge, our work is the first to provide experimental evidence that AA policies have stereotype threat effects.

To investigate whether and to what extent AA policies induce a stereotype threat effect, we conducted a lab experiment with a gender-based quota policy. Participants were assigned to groups of four (two men and two women) and asked to solve quantitative GRE questions; their performance was graded in accordance with the customary practice used in GRE and other standardized admission tests; that is, each correct answer was rewarded while each incorrect answer was penalized. Participants competed for a monetary prize, which was awarded (i) to the top two performers regardless of gender or (ii) to the top two performers subject to a gender quota requiring at least one female winner. To further examine whether AA may serve as a prime generating stereotype threat, we varied whether participants in AA groups were presented with information on the superior average performance of men over women on the actual quantitative GRE exam (informational priming).[2]

We analyze the effect of the gender quota on the performance of men and women by comparing their scores on the exam across the AA conditions, and we examine whether AA has a similar effect with or without informational priming. We find that, while the presence of an AA policy leaves men (of any ability level) unaffected, it changes the performance of women. Women with low baseline ability perform significantly better in the presence of an AA policy and women with high baseline ability perform significantly *worse*. We find the negative effect of AA on the performance of high-ability women and the no effect on men regardless of whether we use linear

[2] Informational priming in this context refers to using objective information as a prime or stimulus of a negative stereotype.

or non-linear specification, or whether we use standard regression analysis or bootstrapping method.

We explore whether the result we get, that women with higher ability underperform under AA, stems from alternation explanations. Namely, high ability women may get annoyed or discouraged by the fact that they get to win mainly because of the fact that they are women, or high ability women may feel less challenged when they compete with other women – they may perceive the competition under AA as a single-sex competition rather than mixed-sex competition.

(this paragraph instead the paragraph in brackets – I got a little bit confused – in the paper we check two things – one related to effort and the other to single sex competition – here you write it as 3 alternative stories but the second and third are actually the same.)

[We explore whether this result, of lower performance by high-ability women under AA, stems from alternative explanations. Namely, whether high ability women reduce their effort in response to AA because they are discouraged that they may win only because they are women, or because they feel the competition is less challenging; or perhaps because they perceive the competition under AA as a single-sex rather than mixed-sex competition. Yet we] We find no evidence that high-ability women reduce their effort in response to the AA policy or that this effect is due to single-sex competition. The results are therefore consistent with the presence of both a competition effect and a stereotype threat effect, and the overall effect can be explained as the net effect of two factors: the increased incentive to compete due to the increased likelihood of winning the prize and the stereotype threat effect. The first factor of greater incentives is likely stronger for the low-ability women as the AA significantly increases their chances to win the prize while the high-ability women already have high chances of winning even without the AA. The increased incentives to perform for lower ability women is sufficient to result in a positive net effect, while the weak change in incentives for the high-ability women (if at all) yields a negative net effect that reveals the existence of an additional negative stereotype threat effect.

Our analysis thus suggests that a comprehensive examination of AA policies should consider the possibility that AA may unintentionally trigger a stereotype threat effect. Although further evidence is needed before drawing any policy conclusions, the findings in this paper do indicate that stereotype threat is relevant to the assessments of AA policies and, consequently, further experimental or empirical work on the subject would be worthwhile.

The remainder of our paper is organized as follows. Section 2 describes our experimental design and procedure. Section 3 describes and discusses our results. Section 4 concludes.

2. Experimental Design and Procedure

To test whether gender-based AA can trigger a stereotype threat effect, we used a between-subject experimental design with random assignment to gender-based AA in a competitive setting with incentivized performance. Participants answered questions from past quantitative GREs[3], a complex task used in actual graduate school admissions. Because men's average performance on such tests is known to be better than women's, we consider this setting appropriate for investigating the existence of a stereotype threat effect.[4] To examine the effect of AA on performance, we calculated a score that penalizes guesses—as is standard practice on exams such as the GRE. Specifically, we awarded a point for each correct answer and subtracted a quarter point for each incorrect answer.

[3] Multiple-choice questions with four answer choices of which the examinee must select one.

[4] We confirm this perception among individuals from the same subject pool by conducting a survey among 75 Harvard students regarding the gender-gap in GRE math performance. Each subject was asked to report his or her belief on women's GRE math score compared to men; specifically, whether women's score is "much lower" (1), "lower" (2), "the same" (3), "higher" (4), or "much higher" (5) than men's. Overall, the Harvard students believe that women's GRE quantitative performance is lower compared to men. This holds true whether they are asked about the general population that takes the GRE (average response is 2.7, significantly different from 3 at the 1% level) or about the Harvard student body (average response is 2.78, significantly differ from 3 at the 1% level).

To check for ability-based heterogeneity in the response to the preferential policy, we designed the experiment in three rounds of 10-minute math exams, where participants were paid according to their performance in each round. In the first round we use noncompetitive piece rate incentives that enable us to use performance in this round as a proxy for ability.

In the second-round exam, which is the main focus of our analysis, subjects were randomly assigned to a group of two men and two women[5] and competed within this group for a bonus of $10 on top of their pay for performance (the pay for performance portion was the same as in the first round). The groups in this round were randomly assigned to one of three conditions: the control condition (No AA), the AA condition (AA), or the AA and informational prime (AA-I) condition. In the No AA condition, the two subjects with the highest scores won the tournament and each received the $10 bonus. In the AA and AA-I conditions, the two subjects with the highest scores each received the $10 bonus, subject to a gender quota that required at least one woman to receive the bonus. That is, if the two highest scores were both earned by men, then the highest scorer and the highest female scorer earned the bonus. The AA-I condition was identical to the AA condition except that participants assigned to the AA-I condition also received an informational stereotype prime prior to the exam. We included this manipulation to compare the effect of the quota policy alone, which may convey information that acts as a stereotype prime, to the effect of a direct stereotype threat prime similar to primes used in previous studies (e.g., Spencer et al. 1999).

The direct prime was included in the description of the quota policy as follows: "*Since ETS statistics show that females quantitative GRE scores are consistently lower compared with males by about 15 percent, we set the following rule: The two participants with the highest score in the group of four (two men, two women) will get the bonus, as long as at least one of the two is a woman. That is, if neither of the participants with one of the top two scores is a woman, the bonus will be given to the participant with the highest score overall, and to the female participant with the highest score. In other words, one of the two winners must be a woman.*"

[5] It was not possible for participants to identify the other members of their group.

Finally, in the third round, subjects were paid according to their scores, as in the first round. After completing the three rounds but before learning what they earned for them and whether they had won the bonus, subjects filled out a questionnaire in which they were asked to self-report their SAT scores (quantitative and verbal), the extent to which they exerted effort on our exam, two Cognitive Reflection Test (CRT) questions used to proxy IQ[6], and their expected score in round 2. They were then informed of their earnings in the three rounds, told whether they had won the $10 prize, and paid privately in cash. Average earnings from performance on the study (all rounds) were $25.43.

The experiment was programmed using Authorware 7.01 and run on computers in the Harvard Decision Science Lab. In total, 248 subjects participated in the study—80 subjects in the control condition, 84 subjects in the AA, and 84 subjects in the AA-I condition—and each condition contained equal numbers of men and women.[7] All subjects were undergraduate or graduate students from Harvard University recruited from the lab's subject pool, and were 20 years old on average. The self-reported average quantitative SAT score was 729.73 and the average verbal SAT score was 719.46.

3. Results

To investigate the effect of implementing gender-based AA policies on performance in a competitive math test and to explore whether AA acts as a stereotype threat prime for women, we focus on the effect of AA on the change in test scores between the first and second rounds. The test score is the appropriate measure to examine, as it determines subjects' payments and captures both the quantity and accuracy of their responses. We analyze the effect of AA on the second-round score using a weighted least squares (WLS) model that adjusts for a systematic

[6] The two questions were: (1) "a bat and a ball cost $1.10 in total. The bat costs a $1 more than the ball. How much does the ball costs?" and (2) "if it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?" These are two out of three questions of The CRT that is thought to be associated with IQ (Frederick 2005).

[7] Two subjects were dropped from our analysis because they left before the end of the experiment.

relationship between variance in score and number of questions attempted.[8] We also explore the effect of AA on the number of questions attempted and answered accurately, measures that may capture changes in effort and response strategy.

3.1. The Net Effect of Affirmative Action

Table 1 presents descriptive information on gender differences in performance for each round. In both the baseline round and the second round, men scored higher on average than women: in rounds 1,men's average score is 6.36 and women's is 5.65; in round 2 men's average score is 7.31 while women's is 6.45. Examining the average number of questions attempted and the average accuracy, we find systematic gender differences that appear to reflect different response strategies: men answered significantly more questions in every round, whereas in two of the three rounds, women were slightly more accurate than men. However, the difference in accuracy is not statistically significant.

Given the observed gender difference in baseline ability, the potential differences in response strategy, even if weak, and the possibly of different effect of AA on men and women, we opt to analyze the effect of AA on the change in score between the first two rounds for each gender separately. In doing so, we control for ability using the first-round scores and for response

[8] A technical concern of using OLS is heteroskedasticity, where the variance in the second-round score may systematically increase with the number of questions attempted. This concern is simply due to the fact that with more questions attempted, the potential high and low scores are more extreme. On the basis of this relationship, we use a WLS model with weights proportional to the number of attempted questions in round 2. We get similar results whether we use the OLS model or bootstrap regression model.

strategy using the number of questions attempted. We also control for age and self-reported SAT scores (verbal and quantitative).[9]

First, we examine the effect of the AA policy by pooling both conditions with AA together, whether or not informational priming was provided. We consider two sets of specifications: one with a dummy variable that equals 1 when the AA policy is in effect and zero otherwise, and the other that also includes an interaction term of the AA dummy variable with baseline ability, as measured by the first-round score. The first specification assumes a similar effect of AA for all participants, while the second allows for heterogeneity of the response to AA by ability. The second specification is important since there are several reasons to suspect that the policy may affect low and high ability individuals differently; for instance, low ability men may give up under AA while low ability women's performance may get the greatest boost due to their increased opportunity to win.

Table 2 reports the results; as expected, and regardless of the specification considered, baseline ability (first-round score) has a negative effect on the change in scores between the first two rounds and is highly significant for women. This negative effect is expected since it is more difficult to improve when starting off from a higher baseline score. We also find that the self-reported quantitative SAT score has a positive significant effect on the change in scores for both men and women; however, the self-reported verbal SAT score does not. The number of questions attempted in the first round, which may capture response strategy,[10] is associated with a significantly higher change in scores for women (effect of 0.442 or 0.471, depending on the specification, both with p-value < 0.001). The effect for men is less than half that for women and is not significant.

---

[9] There is no significant gender difference in average quantitative SAT score in our sample: women's average quantitative SAT score is 736 while men's is 718 (one-sided *t*-test yield $p = 0.11$).

[10] Indeed, in a regression of the success rate in round 2 on the number of attempts, gender, and their interaction, we find that for a given number of attempts, women have significantly higher success rates (the main effect of gender is 0.168, $p = 0.054$), which diminish marginally with the number of attempts (the interaction is –.012 $p = 0.107$). Nevertheless, the gender effect is positive up to 14 attempts, representing about 92% of the women in our sample. This is consistent with a different response strategy across gender, where men attempt more questions at the cost of accuracy.

Turning to the main question—examining the effect of AA on performance—we find that, regardless of ability (Table 2, column 1), AA has a positive albeit insignificant effect on women. When we allow for the possibility that high- and low-performing women are affected differently (Table 2, column 2), however, we find a positive and significant main effect of AA (1.889, p-value = 0.019), which declines with ability (–0.23; p-value = 0.035). With respect to the average woman in our sample, the overall effect of AA is about 10 percent of the score in round 1 ((1.889-0.23*5.65)/5.65). However, the overall effect of AA becomes negative for women whose first-round score is over 8.21. Looking at percentiles, approximately 80% of women have a first-round score lower than 8 in round 1, which is very close to this cutoff. In sum, AA has a positive effect on most women, but the projected effect of the policy is negative for the top 20%.[11,12] For men, there is an insignificant positive effect of AA whether or not we allow the effect to vary with ability. Specifically, the main effect of AA on men's performance, regardless of ability, is 0.566 and is insignificant (Table 2, column 3). When AA is allowed to have a differential effect by ability, its main effect is positive (1.051) and its interaction with ability is negative (-0.067), both are highly insignificant.

To test whether the overall negative effect on the high-ability women is significant, and since sample size becomes an issue when examining this subgroup, we use the bootstrap method. For this exercise, we split the women in our sample into three groups according to their first round scores—low ability (scores below 5), mid ability (scores between 5 and 8), and high ability (scores of 8 or higher). The brackets were selected such that the low-ability group includes women whose round 1 score is below the median, then we split the upper half into the mid-ability group which includes women whose round 1 score is between 50%-75%, and the high-ability group with women whose round 1 score is in the top 25%. The high ability group corresponds closely to the

[11] Note that since we are working with a sample of students from a very selective university, the results for the lower range of the ability distribution in our sample may be more representative of a broader population than are the results for the entire sample. At the same time, AA policies often aim at the very best individuals within the beneficiary group, in which case the very best in our sample would be the most interesting and relevant individuals to examine.

[12] Examining the correlation between the score in round 1 and the score in round 2, as well as between the number of questions attempted across the two rounds, we find very high correlation of about 0.8 for both. That is, the effect is not driven by high ability women who performed poorly in the first stage.

group for whom the overall effect of AA appears to be negative. We then calculate for each subgroup the difference in (mean) scores between round 2 and round 1, we calculate it separately for women who were under AA and women who were in the control group, and then take the difference-in-differences (AA minus control). We repeat this exercise 500 times, randomly sampling women in each subgroup with repetition, resulting in a distribution of difference-in-differences (see Figure 1). This allows us to ask whether the average diff-in-diff is negative and significant for the subgroup of high-ability women. We find that the mean diff-in-diff for the high-ability women is negative in 93% of the random sample with mean equal to −1.35, slightly positive but insignificant for the mid-ability women with mean equal to 0.156, and positive for the low-ability women in 85% of the random sample with mean equal to 0.677.

The advantage of the bootstrapping method is that it doesn't require any parametric assumptions. That is, the effect is seen even without any assumption on the functional form. Nevertheless, we further complement this exercise with a regression analysis by grouping women according to the three ability groups described above; this exercise allows for non-linear relationship between ability and the effect of AA. The results (see Table 3) are consistent with those obtained from the bootstrap approach: the effect of AA on the low-ability women is positive (1.43) and significant (one-sided t-test yields $p = 0.03$); the effect of AA on mid-ability women is overall negative (−0.118) but insignificant, and the effect of AA on high-ability women is negative (−1.67) and significant (one-sided t-test yields $p = 0.03$.)

The positive main effect of AA on women's scores reported in Tables 2 and 3 is not surprising given that the gender quota increases (at least weakly) a woman's objective chance to win the bonus. The finding that the positive effect of AA decreases with ability—to the point that it has a negative effect on high-ability women—is surprising and suggests that there is another factor offsetting the main effect.

Before moving on to explore whether the results are due to stereotype threat effect, one may wonder whether the results are robust to using a different measure of ability that is not directly related to the performance in the study. To explore this issue, we constructed groups based on an index composed of (1) reported SAT quantitative scores, and (2) the two CRT questions which can be thought of as a proxy for IQ.[13] The combined index takes the two measures into account, as well as the speed of response to the IQ questions. All the results reported in the paper are robust to splitting participants into high, mid, and low ability groups according to this index rather than round 1's score. Nevertheless, we believe that using the score in round 1 is the most appropriate proxy for ability in this study exactly because it is directly relevant for the task in round 2, it is observed rather than self-reported, and is measured at the time subjects take the test. The SAT score, in contrast, is self-reported, taken at least a year before the study took place, and may have been taken multiple times in the attempt to get the best possible score. The CRT questions, although positively correlated with SAT scores (with a correlation of 0.4), do not necessarily measure the same skills, and some individuals may already be familiar with these questions prior to participating in the study.

3.2. Does Affirmative Action Act as a Prime?

Can the observed negative effect of AA be due to AA acting as a negative prime? To address this question, we first examine whether the effect observed is driven by women exposed to the informational prime or could be obtained by AA alone.

Table 4 presents the results with a specification similar to Table 3 but with an additional indicator variable "Info," which takes the value of 1 if a participant was assigned to the AA treatment with the additional informational prime and 0 otherwise.[14] We also interact this indicator variable with our measure of ability and find that the Info main effect and its interaction with ability group are insignificant. This suggests that the effect of AA on women is the same whether or not participants receive a direct informational prime, and is consistent with the idea that AA by itself is acting as a prime.

Using the results reported in Table 4 and the distribution of first-round scores for women, we find that women in the first two ability groups with first-round scores below 8—about 80% of

[13] The correlation between this measure of ability and the one based on first round score is 0.63.

[14] The information prime describes women's inferior performance relative to men on GRE, which is, on average, 15% lower. See Section 2 for details.

the women in our sample—improved or had a similar change in scores under the AA treatment compared to the control. By contrast, women in the high-ability group—the top 20%—performed worse under AA. The results for men shown in Table 4 are very similar to those shown in Table 3, indicating that adding the informational prime dummy does not change the results.

3.2.1. Two Alternative Explanations

Finding no effect of the informational prime beyond AA is consistent with AA being a prime that triggers a stereotype threat. However, it is also possible that the gender stereotype has no effect at all—that is, that AA encourages low-ability women by sufficiently increasing the marginal benefit of their effort while at the same time making the marginal benefit of extra effort not worth it for the high-ability women. This could explain the observed pattern in the effect of AA: positive for low-ability women and negative for high-ability women. Another possibility is that AA leads women to focus on single-sex competition, which may encourage low-ability women to compete, consistent with the finding that women compete more in a single-sex competition (e.g., Gneezy et al. 2003), while at the same time reducing high-ability women's concern about the competition. We examine these alternative explanations next.

Effort Response

To test whether women, particularly high-ability women, reduced their effort in response to AA we look at several measures. We start by examining participants' self-reported effort exerted during the study, as indicated in an exit questionnaire (responses range from 1 to 7, where 7 represents the highest effort). Table 5A and 5B report the results of ordered probit regressions of those responses on the same specification used in Table 2; namely, AA dummy variable, first-round score, and their interaction, plus controls that include the number of questions attempted in the first round, age, and self-reported SAT scores (quantitative and verbal). Table 5A presents the results for women and Table 5B for men. Both tables report results from the two specifications considered before: one using the continuous measure of ability forcing a linear relationship and the second using ability groups allows for nonlinear relationship. We also note that only three individuals (two women and one man) reported a low effort level of 1 or 2. The tables, therefore, present the results both including these individuals (columns 1, 3, and 5) and excluding them (columns 2, 4, and 6). The results show that women report exerting more effort under AA, and this finding is robust to adding the interaction of the dummy variable of quota policy with ability (measured by first-round score) when the outliers are excluded. Including all women, including the two outliers, the effect is always positive albeit sometimes insignificant. For men, we find an insignificant effect of AA, whether or not the effect of AA is allowed to vary with ability and the outlier is excluded. Hence, even if AA changes women's perception of their probability of winning, or that it discourages them because under AA they would think they won just because they are woman, it does not decrease their self-reported effort in response; in fact, the evidence suggests that they increase their effort across the board.[15]

Next, we examine the number of question attempted, which is often used as a crude proxy of effort (Fryer et al 2008). The idea is that with more effort, individuals attempt to solve more questions. The results are in Table 6, and show insignificant main effect of AA for both men and women, as well as no relationship between ability, AA, and number of questions attempted.

In the exit questionnaire we also asked participants what they think their *score* in round 2 was. This was, of course, before they received any feedback on their actual performance in the study. If high ability women exerted less effort under AA, they should report lower scores in round 2 compared to high ability women in the control condition (without AA). Examining the responses, we find that overall women under AA reported higher scores in round 2 (average predicted score is 6.2) compared to those in the control (with average predicted score of 5.8). This positive difference is not statistically significant, however, examining the high-ability women separately, we find large positive difference of 2.25 points on average that is significant at the 5% level. Comparing the men's predicted score in round 2 under AA vs. the control, we find no difference overall or by ability.

[15] Given that the effort measure is self-reported, it may be affected by the condition. In this case, we can interpret the results to suggest that women do not believe they should exert less effort under AA.

We complement the self-reported measures and the number of questions attempted by a back-of-the-envelope calculation to examine whether it was optimal for high-ability women to reduce their effort under AA compared to the control. Formally, each individual's objective function is choose an effort level to maximize $p(e; AA)Bonus + Reward(e) - c(e)$, where e is effort, and $p(e;AA)$ is the probability of winning the bonus as a function of effort e and the AA environment (AA=1 represents AA policy). *Reward*(e) is the piece-rate earning as a function of effort, and $c(e)$ is the cost of exerting effort level e. Assuming increasing cost of effort, greater effort under AA is expected if $p'(e; AA = 0) < p'(e; AA = 1)$.

We therefore calculate the marginal effect of effort on the probability of winning for high-ability women (round 1 score of 8 or above), once under AA policy and once without AA.[16] We then take the individual difference in the marginal effect (AA minus no-AA) and find that the average difference across high ability women is positive, meaning that the marginal effect of effort on the probability of winning for high-ability women is higher—rather than lower—under AA compared with the control (the average is 0.0193, p<0.01.)

Although this is a back-of-the-envelope calculation, it nevertheless provides evidence that it was not optimal for high-ability women to reduce their effort in response to AA. Indeed, the women in our study did not seem to do that. Taken together, this suggests that a reduction in effort does not explain the lowered performance of high-ability women under AA.

---

[16] Specifically, we took each woman's score in round 1 and calculated whether or not she would win if she kept the same level of effort (same score) and given a randomly assigned group's round 2 scores. We repeated this exercise assuming she solved one more question correctly (as a proxy for effort), and take the difference. This is the marginal effect of effort on the probability of winning with that particular group. For the calculation of marginal probability under AA, we assign each woman 500 random groups that were subject to AA in the second round, calculate her marginal probability with each group, and take the average. This is her marginal effect of effort on probability of winning with AA. For the marginal effect on the probability of winning without AA, we do a similar calculation, assigning each woman to 500 random groups who were not subject to the AA policy in round 2. Note that the marginal winning probability is calculated by assuming an additional question solved correctly, since it is similar to the actual average difference in score between round 1 and round 2. That is, the marginal winning probabilities calculated are relevant for the effort decision made in round 2.

Women-Only Competition

To look into the possibility that single-sex competition reduces high-ability women's concern and effort while encouraging low-ability women to compete, we run an additional condition in which the second-round competition is between two women. Thirty-four women from the same subject pool participated in this condition[17]. The setting for this women-only condition was identical to the one described above. That is, participants were asked to solve GRE quantitative questions in three rounds, and as in the other conditions, in round 1 and round 3 they were paid according to their performance (piece rate for net performance). The difference was in round 2 where participants in the women-only conditions were assigned to single-sex groups of 2 (two women), where the person to win the bonus is the woman with the highest score.

Using the data from this condition, we test whether women's performance in the paired, women-only condition differs from women's performance in the control group, where subjects are assigned to mixed-sex groups and are not subject to AA. Table 7 shows no significant difference between the two groups: in both conditions, neither the main effect nor its interaction with ability is significant. In other words, the effect of AA on women's performance does not seem to be due to a shift in focus from a mixed-sex group to a single-sex group.

Taken together, the findings are inconsistent with higher-ability women exerting less effort in response to AA due to their focus on competing against another woman or because they have higher chances of winning. Hence, the negative effect of AA on high-ability women supports the hypothesis that the AA policy evokes a stereotype-threat effect. It is interesting to note that the result—that the negative stereotype threat effect is evident only among high-ability women—is consistent with Preckel et al. (2008), who found evidence that the greatest gender gap in confidence in one's ability is present among the most gifted children. This is also consistent with the stereotype threat literature, which suggests that the stereotype effect presents itself among those who highly identify with the domain of the task (e.g., Aronson et al. 1999, Steele et al. 2002,

[17] In sessions in which we ran this additional condition, we also ran the other conditions. That is, subjects were randomly assigned to either this women-only condition or to one of the previous conditions with groups of four.

Wei 2009). In our study, high first-round ability may be a proxy for identification with the math domain. If this is the case, the stereotype threat theory would predict that the effect will be concentrated among the high-ability women.

4. Discussion and Conclusion

We find that AA affects the performance of women on quantitative GRE questions in an incentivized and competitive environment. We observe a positive and significant effect on women with lower baseline ability but a negative and significant effect on women with higher baseline ability. Calculating the marginal return to effort across conditions reveals that reducing effort is not optimal for high ability women, and self-reported effort among women, as well as their predicted score or number of question attempted, all confirm that AA does not lead to an intentional effort reduction. On the contrary, women report higher effort, and high ability women believe that their score is higher under AA relative to no-AA. We also find no evidence that a single-sex competition between women explains the negative effect on the performance of high-ability women. Instead, the negative effect of AA on performance of high ability women stems from a reduction in their success rate, i.e., AA reduces the effectiveness of their effort (see table A1 in the Appendix). This is especially relevant since we find that AA in and of itself has a similar effect on performance whether or not it is accompanied by a direct stereotype prime, and we confirm that Harvard students believe women perform worse than men on the GRE quantitative exam. These results, together, are consistent with the hypothesis that AA acts as a stereotype prime that leads to the unintended negative consequence of impairing performance of the protected group.

Moreover, the results in this study may be underestimated, as women in the control group were asked to solve complex math questions in a clearly mixed gender competition; psychologists may argue this environment already creates stereotype threat even absent AA. If this is indeed the case, the results reported in this study are merely the effect of an incremental change in stereotype threat across the control and the AA conditions.

As noted in the paper, participants in the study are students in a highly selective institution, and therefore may not be representative of the effect of AA in the field. There are two points to raise in regard to this objection: first, AA policies are often designed to motivate candidates of the highest abilities, and therefore the results in this study are relevant. Second, this study makes the point that stereotype threat may be triggered with well-intentioned policies, and this should be taken into consideration and be further investigated. The fact that we find this even among women of such high ability is important to note, and frankly, worrisome. Yet it is possible that the effect is driven by relative ranking within the group rather than absolute skill. If this is true, then the effect found in this study may also occur in other groups of lower abilities, as long as the high ability individuals are defined relative to the group they are competing against.

With the substantial interest around the world in AA policies aimed at advancing women, it is important to understand fully the potential effect of such policies. This paper provides first evidence that AA policies can trigger stereotype threat effects. We hope that future work will examine this issue further to provide additional evidence on the prevalence and magnitude of the stereotype threat effect of AA polices.

References

Aronson, Joshua, Michael J. Lustina, Catherine Good, and Kelli Keough. 1999. When white men can't do math: necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology* 35: 29-46.

Balafoutas, Loukas, and Matthias Sutter. 2012. Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science* 335:579–582.

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

Catalyst Census. 2013. Fortune 500 Women Board Directors. Available at: http://www.catalyst.org/system/files/2013_catalyst_census_fortune_500_women_board_director.pdf.

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

European Commission – Directorate-General for Justice, 2012. Women in Economic Decision-making in the EU: Progress Report.
Available at: http://ec.europa.eu/justice/gender-equality/files/women-on-boards_en.pdf.

Frederick Shane. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspective* 19(4): 25-42.

Fryer, Roland, Steven Levitt, John List. 2008. Exploring the impact of financial incentives on stereotype threat. *American Economic Review: Papers and Proceedings* 98(2):370-375.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. Performance in competitive environments: Gender differences. *Quarterly Journal of Economics* 118:1049–1074.

Hyde, Janet Shibley, Elizabeth Fennema, and Susan J. Lamon. 1990. Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin* 107(2):139–155.

Meeting of the OECD Council at Ministerial level, Paris 25-26 May 2011, Report of the Gender Initiative: Gender Equality in Education, Employment and Entrepreneurship. Available at: http://www.oecd.org/education/48111145.pdf.

National Association for Law Placement press release, December 11, 2013. Representation of Women Associates Falls for Fourth Straight Yeas as Minority Associates Continue to Make Gains – Women and Minority Partners Continue to Make Small Gains.

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

Niederle, Muriel, and Lise Vesterlund. 2010. Explaining the Gender Gap in Math Test Scores: The Role of Competition. *The Journal of Economic Perspectives* 24(2): 129-144.

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

Preckel Franzis, Thomas Goetz, Reinhard Pekrun, and Michael Kleine. 2008. Gender Differences in Gifted and Average-Ability Students: Comparing Girls' and Boys' Achievement, Self-Concept, Interest, and Motivation in Mathematics. *Gifted Child Quarterly* 52(2): 146-59.

Schotter, Andrew, and Keith Weigelt. 1992. Asymmetric tournaments, equal opportunity laws, and affirmative action: Some experimental results. *Quarterly Journal of Economics* 107(2):511–539.

Shih, Margaret, Todd L. Pittinsky, and Nalini Ambady. 1999. Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science* 10:80–83.

Spencer, Steven J., Claude M. Steele, Diane M. Quinn. 1999. Stereotype threat and women's math performance. *Journal of Experimental Social Psychology* 35:4–28.

Steele, Claude M. and Joshua Aronson. 1995. Contending with a stereotype: African-American intellectual test performance and stereotype threat. *Journal of Personality and Social Psychology* 69:797–811.

Steele, Claude M., Steven J. Spencer, and Joshua Aronson. 2002. Contending With Group Image: The Psychology of Stereotype and Social Identity Threat. *Advances in Experimental Social Psychology* 34: 379-440.

Stricker, Lawrence J. 1998. Inquiring about examinees' ethnicity and sex: Effects on AP Calculus AB examination performance. Collage Board Report No. 98-1. ETS Report No. 98-5.

Sutter, Matthias and Daniela Rützler. 2010. Gender Differences in Competition Emerge Early in Life. IZA Discussion Paper No. 5015.

U.S. Department of Commerce, Economics and Statistics Administration. 2011. Women in STEM: A Gender Gap to Innovation.
Available at: http://www.esa.doc.gov/Reports/women-stem-gender-gap-innovation.

Wei, Thomas. 2009. Under what conditions? Stereotype threat and prime attributes. Working Paper. Available at http://www.people.fas.harvard.edu/~twei/papers/sthreat_exper.pdf.

Wei, Thomas. 2012. Sticks, stones, words, and broken bones: New field and lab evidence on stereotype threat. *Educational Evaluation & Policy Analysis* 34:465.

Table 1: Summary Statistics

		Male (mean)	Female (mean)	Diff.
Round 1	Score	6.36	5.65	p<0.10 one sided
	# Questions	8.91	8.32	p<0.10 one sided
	Ratio correct	0.76	0.72	insignificant
Round 2	Score	7.31	6.45	p<0.10 two-sided
	# Questions	11.50	9.89	p<0.01 two-sided
	Ratio correct	0.69	0.71	insignificant
Round 3	Score	7.54	7.06	insignificant
	# Questions	12.20	11.26	p<0.05 two-sided
	Ratio correct	0.68	0.68	insignificant

* p<0.1, ** p<0.05, *** p<0.01

Table 2: The Effect of Affirmative Action on Performance

	(1) Female	(2) Female	(3) Male	(4) Male
Affirmative Action (AA)	0.436	1.889**	0.566	1.051
	(0.432)	(0.795)	(0.532)	(0.904)
Score in 1st round (Score R1)	-0.723***	-0.572***	-0.372***	-0.340**
	(0.078)	(0.093)	(0.141)	(0.137)
(AA)x(Score R1)		-0.230**		-0.067
		(0.108)		(0.128)
# Questions in 1st round	0.442***	0.471***	0.187	0.171
	(0.069)	(0.075)	(0.132)	(0.130)
Age	0.000	-0.052	-0.173	-0.160
	(0.106)	(0.111)	(0.180)	(0.177)
SAT Quantitative	0.018***	0.017***	0.022***	0.023***
	(0.003)	(0.003)	(0.006)	(0.006)
SAT Verbal	0.002	0.002	0.002	0.002
	(0.003)	(0.003)	(0.004)	(0.004)
Constant	-13.242***	-13.153***	-13.356**	-14.058**
	(3.974)	(3.931)	(5.905)	(6.019)
N	123.000	123.000	118.000	118.000
r2	0.390	0.416	0.218	0.221

Notes: WLS regressions. Dependent variable: difference in score between round 2 and round 1. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.

Table 3: The Effect of Affirmative Action on Performance, by Group

	(1) Female	(2) Male
Affirmative Action (AA)	1.434*	0.979
	(0.744)	(0.705)
1st Rnd Score Group	-1.071*	-1.270*
	(0.582)	(0.674)
AA x 1st Rnd Score Group	-1.552**	0.108
	(0.670)	(0.637)
# Questions in 1st round	0.195**	0.013
	(0.095)	(0.082)
Age	-0.018	-0.123
	(0.140)	(0.177)
SAT Quantitative	0.012***	0.022***
	(0.004)	(0.006)
SAT Verbal	-0.001	0.001
	(0.004)	(0.004)
Constant	-7.887*	-13.626**
	(4.578)	(6.313)
N	123	118
r2	0.229	0.203
T_Low_pos	0.028	0.084
T_High_neg	0.034	0.901

Notes: WLS regressions. Dependent variable: difference in score between round 2 and round 1. T-Low(+) displays the result of a t-test calculating the probability that an individual in the lower performance group was positively impacted by the Affirmative Action. T-High(-) displays the result of a t-test calculating the probability that an individual in the higher performance group was negatively impacted by the Affirmative Action. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.

Table 4: Affirmative Action and Informational Prime

	(1) Female	(2) Male
Affirmative Action (AA)	1.696**	1.350t
	(0.755)	(0.830)
Info	-0.519	-0.616
	(0.701)	(0.958)
1st Rnd Score Group	-1.071*	-1.277*
	(0.586)	(0.683)
AA x 1st Rnd Score Group	-1.601**	-0.048
	(0.741)	(0.784)
Info x 1st Rnd Score Group	0.064	0.257
	(0.753)	(0.842)
# Questions in 1st round	0.196**	0.021
	(0.096)	(0.082)
Age	-0.026	-0.135
	(0.135)	(0.181)
SAT Quantitative	0.012***	0.021***
	(0.004)	(0.006)
SAT Verbal	-0.001	0.001
	(0.004)	(0.004)
Constant	-7.644*	-12.898*
	(4.526)	(6.811)
N	123	118
r2	0.234	0.207

Notes: WLS regressions. Dependent variable: difference in score between round 2 and round 1. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.

Table 5A: The Effect of Affirmative Action on Effort, Females

	(1)	(2)	(3)	(4)	(5)	(6)
Affirmative Action (AA)	0.405*	0.473**	0.549t	0.878**	0.497t	0.668*
	(0.211)	(0.221)	(0.374)	(0.404)	(0.318)	(0.353)
Score in 1st round (Score R1)	0.036	0.041	0.052	0.086t		
	(0.049)	(0.049)	(0.052)	(0.057)		
(AA)x(Score R1)			-0.026	-0.071		
			(0.051)	(0.055)		
1st Rnd Score Group					0.187	0.392t
					(0.247)	(0.270)
AA x 1st Rnd Score Group					-0.093	-0.200
					(0.256)	(0.279)
# Questions in 1st round	-0.017	-0.005	-0.014	0.006	-0.007	-0.002
	(0.048)	(0.046)	(0.049)	(0.045)	(0.037)	(0.037)
Age	0.017	0.009	0.012	-0.006	0.012	-0.001
	(0.064)	(0.068)	(0.067)	(0.073)	(0.066)	(0.071)
SAT Quantitative	0.005**	0.002	0.005**	0.002	0.005**	0.002
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
SAT Verbal	-0.002	-0.003	-0.002	-0.003	-0.002	-0.003t
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	123	121	123	121	123	121

Notes: Ordered probit regressions. Dependent variable: self-reported effort level. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.

Table 5B: The Effect of Affirmative Action on Effort, Males

	(1)	(2)	(3)	(4)	(5)	(6)
Affirmative Action (AA)	0.083	0.142	0.205	0.293	0.082	0.152
	(0.198)	(0.202)	(0.420)	(0.455)	(0.272)	(0.283)
Score in 1st round (Score R1)	0.129**	0.147***	0.139**	0.159**		
	(0.054)	(0.056)	(0.064)	(0.068)		
(AA)x(Score R1)			-0.019	-0.023		
			(0.063)	(0.067)		
1st Rnd Score Group					0.349*	0.368*
					(0.206)	(0.212)
AA x 1st Rnd Score Group					-0.160	-0.204
					(0.233)	(0.239)
# Questions in 1st round	-0.094*	-0.112**	-0.097**	-0.116**	-0.030	-0.034
	(0.050)	(0.053)	(0.049)	(0.052)	(0.043)	(0.045)
Age	0.005	0.000	0.008	0.003	-0.011	-0.017
	(0.063)	(0.065)	(0.063)	(0.065)	(0.065)	(0.066)
SAT Quantitative	0.003	0.002	0.003	0.002	0.004*	0.004t
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
SAT Verbal	0.000	0.001	0.000	0.001	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	118	117	118	117	118	117

Notes: Ordered probit regressions. Dependent variable: self-reported effort level. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.

Table 6: Number of Questions

	(1) Female	(2) Female	(3) Male	(4) Male
Affirmative Action (AA)	-0.529	-0.085	1.211	0.631
	(0.904)	(0.673)	(0.911)	(0.778)
Score in 1st round (Score R1)	0.178		0.085	
	(0.149)		(0.143)	
(AA)x(Score R1)	-0.037		-0.064	
	(0.138)		(0.110)	
1st Rnd Score Group		1.252*		-0.364
		(0.654)		(0.662)
AA x 1st Rnd Score Group		-0.570		0.015
		(0.711)		(0.636)
# Questions in 1st round	0.526***	0.545***	0.632***	0.723***
	(0.197)	(0.148)	(0.150)	(0.085)
Age	0.091	0.074	-0.083	-0.088
	(0.249)	(0.244)	(0.150)	(0.143)
SAT Quantitative	0.009*	0.008**	0.010**	0.013**
	(0.005)	(0.004)	(0.005)	(0.005)
SAT Verbal	-0.005t	-0.006t	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
N	123	123	118	118
r2	0.600	0.612	0.674	0.675

Notes: WLS regressions. Dependent variable: number of questions answered in round 2. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.

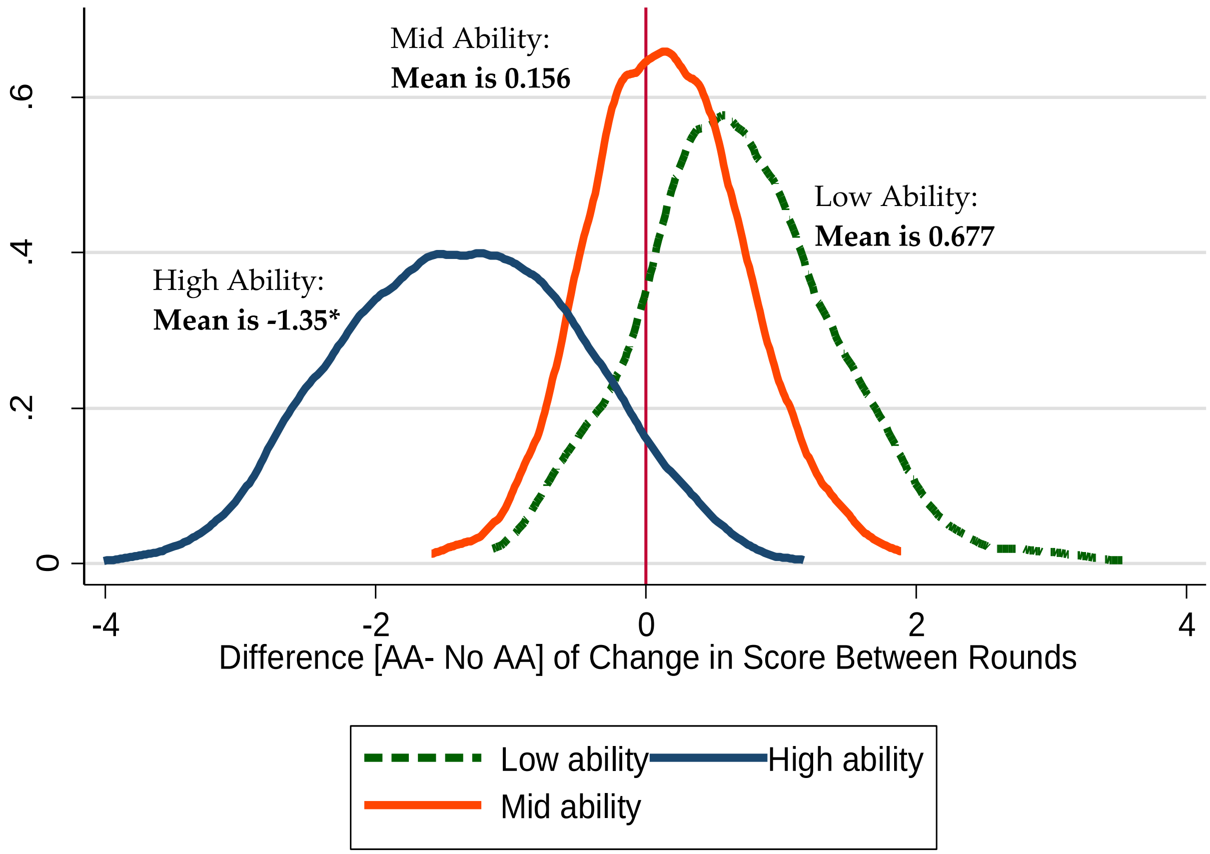
Table 7: Women-Only Competition

	(1)	(2)
Women only	0.235	0.366
	(0.802)	(0.743)
Score in 1st round (Score R1)	-0.451***	
	(0.124)	
1st Rnd Score Group		-0.667
		(0.586)
(Women only)x(Score R1)	0.025	
	(0.095)	
(Women only)*(1st Rnd Score Group)		-0.248
		(0.592)
# Questions in 1st round	0.311**	0.074
	(0.122)	(0.078)
Age	-0.066	-0.082
	(0.120)	(0.146)
SAT Quantitative	0.012*	0.004
	(0.006)	(0.007)
SAT Verbal	0.003	0.001
	(0.003)	(0.004)
N	72	72
r2	0.193	0.057

Notes: WLS regressions. Dependent variable: difference in score between round 2 and round 1. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.

Figure 1

Females' [Round 2- Round 1] Score Difference, by AA



Appendix

Table A1: Success Rate

	(1) Female	(2) Female	(3) Male	(4) Male
Affirmative Action (AA)	0.149**	0.120**	0.019	0.024
	(0.064)	(0.053)	(0.063)	(0.051)
Score in 1st round (Score R1)	0.019**		0.034***	
	(0.009)		(0.008)	
(AA)x(Score R1)	-0.013*		0.000	
	(0.008)		(0.007)	
1st Rnd Score Group		0.064t		0.115***
		(0.040)		(0.043)
AA x 1st Rnd Score Group		-0.051		-0.045
		(0.043)		(0.040)
# Questions in 1st round	-0.001	0.002	-0.019**	-0.004
	(0.009)	(0.006)	(0.007)	(0.006)
Age	0.004	0.004	-0.013	-0.018t
	(0.010)	(0.010)	(0.010)	(0.011)
SAT Quantitative	0.001***	0.002***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
SAT Verbal	0.001**	0.001**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
N	123	123	118	118
r2	0.416	0.409	0.471	0.419

Notes: OLS regressions. Dependent variable: success rate in round 2. Robust standard errors are reported in parentheses. * indicates significance at the 10-percent level; ** indicates significance at the 5-percent level; *** indicates significance at the 1-percent level. t indicates p<0.15.