The Incentive Effects of Affirmative Action in a Real-Effort Tournament*

Caterina Calsamiglia* — U. Autònoma de Barcelona
Jörg Franke[1] — TU Dortmund
Pedro Rey-Biel** — U. Autònoma de Barcelona

First draft: June 17, 2009
This version: January 14, 2011

Abstract

Affirmative action policies bias tournament rules in order to provide equal opportunities to a group of competitors who have a disadvantage they cannot be held responsible for. Its implementation affects the underlying incentive structure which might induce lower performance by participants, and additionally result in a selected pool of tournament winners that is less efficient. In this paper, we study the empirical validity of such concerns in the case where the disadvantage affects the capacity to compete. We conduct real-effort tournaments between pairs of children from two similar schools who systematically differ in how much training they received ex-ante in school on the task at hand. Our results show that the implementation of affirmative action enhanced the performance of most advantaged or disadvantaged subjects. Additionally, while affirmative action balanced the proportion of disadvantaged individuals winning their respective tournament, the average performance of the pool of winners only decreased slightly.

Keywords: Affirmative action, tournament, real-effort, experiment, sudoku.

JEL classification: C72; C91; J78; M52

---

* We are especially grateful to Antoni Calvó-Armengol for his support and encouragement on this project. We thank Jose Apesteguia, Miguel A. Ballester, Carmen Beviá, Jeanette Brosig, Antonio Cabrales, Marta García-Matos, Maureen Gleeson, Uri Gneezy, Guillaume Haeringer, Nagore Iriberri, Inés Macho-Stadler, Muriel Niederle, Carmit Segal, Tom Palfrey, Neslihan Uler and seminar audiences at many seminars and conferences. We are deeply grateful to the directors, faculty and administrative personnel at Agora, Aula, Colegio Alemán, Emili Juncadella, Pegaso and Sagrado Corazón for their understanding, permission and help with the experiments. We thank Pau Balart, Julen Berasaluce, Ignacio Fernández, Markus Kinateder, Tomasso Majer, Luca Merlino and Natalia Montinari for their help in running the experiments. Caterina Calsamiglia acknowledges financial support by a Ramón y Cajal contract of the Spanish Ministerio de Ciencia y Tecnología, through the Spanish Plan Nacional I+D+I (SEJ2005-01481, SEJ2005-01690 and FEDER), and through the "Grupo Consolidado de tipo C" (ECO2008-04756), the Generalitat de Catalunya (SGR2005-00626 and the Barcelona Economics Program of XREA), and the Consolider-Ingenio 2010 (CSD2006-00016) program. Jörg Franke acknowledges financial support from Ministerio de Educación y Ciencia (SEJ2005-01481/ECON) and FEDER, from Generalitat de Catalunya (2005SGR00454) and from Universitat Autònoma de Barcelona. Pedro Rey-Biel acknowledges financial support from Ministerio de Ciencia y Tecnología (ECON2009-07616 and Consolider-Ingenio CSD2006-00016), Barcelona GSE Research Network and of the Government of Catalonia (2009SGR-169).
* Caterina Calsamiglia. Universitat Autònoma de Barcelona. Department d´Economia i d´Historia Econòmica. 08193, Bellaterra. Barcelona (Spain). E-mail: caterina.calsamiglia@uab.cat.
[1] Jörg Franke, Department of Economics and Social Science, TU Dortmund, 44227 Dortmund, Germany. Tel.: (+49) 231 755 3246. E-mail: joerg.franke@tu-dortmund.de.
** Pedro Rey-Biel. Universitat Autònoma de Barcelona. Department d´Economia i d´Historia Econòmica. 08193, Bellaterra. Barcelona (Spain). Tel: (+34) 935812113. E-mail: pedro.rey@uab.es.

1. Introduction

In many selection processes such as university admissions, job promotions and procurement auctions, competition helps identifying the highest-ability candidates, facilitates the efficient allocation of talent, and provides incentives for individuals to improve their performance. This objective may not be achieved if some otherwise competitive candidates do not stand a fair chance to win the competition. For example, talented students from poor economic backgrounds may have attended schools that receive less funding which, irrespective of their talent, may affect their SAT performance and hence their university admission. Likewise, some individuals may belong to groups which historically have suffered discrimination, and have to overcome major obstacles in order to be on an equal footing to compete.

Affirmative Action (AA) policies have two main objectives: to guarantee that positions are fairly allocated in society and to allow for the creation and identification of talent. AA policies take proactive steps to provide equal opportunities to discriminated groups that have a potential disadvantage.[2] They are often implemented by biasing tournament rules in order to increase the probability of success of an otherwise disadvantaged group. For example, a fixed lump-sum bonus of 20 (out of 150) points was added to the score of minority applicants to the undergraduate program at the University of Michigan and a similar but "unofficial lift" scheme is used at several top universities.[3] In a different domain, public procurement auctions, bid preferences are granted in a multiplicative way. For example, road construction contracts in California are auctioned off by granting a 5% reduction of the submitted bid to small business enterprises.

In this paper we analyze how AA rules can affect individuals' choice of effort, and consequently shape the exhibited distribution of performance by participants in the special case where the disadvantage affects performance directly. As reviewed later in the introduction,

---

[2] Merriam-Webster Online defines affirmative action as "an active effort to promote the rights or progress of minority groups or other disadvantaged persons".

[3] This procedure was recently ruled to be unconstitutional by the Supreme Court, due in part to alleged distortionary effects on incentives that such compensation may create. State funded universities such as California, Florida and Texas have also applied similar policies in the past. A number of papers have analyzed the effects of such banning on the efficiency of race-blind policies that result from colleges preferences for diversity, See, among others, Chan and Eyster (2003) and Fryer and Loury (forthcoming). These papers show that banning affirmative action forces colleges with a taste for diversity to target minorities indirectly through randomly admitting students with relatively lower test scores or other characteristics correlated with race, which leads to substantial inefficiencies.

theoretical literature suggests that in competitive environments performance is maximized if the probabilities of winning depend equally on individual effort across different agents or, in other words, if the competition is symmetric. Hence, the use of unbiased rules is optimal if competitors are equally productive in the respective task, which corresponds to an equal treatment of equals approach. However, in contexts where the disadvantaged are less productive, biasing the competition rules in order to level the playing field changes the incentives of all participants and therefore their performance, which could in turn lead to a more representative and not necessarily worse performing selection of individuals ex-post.

Our experimental study addresses this issue through a series of real-effort tournaments where pairs of children from two different schools compete against each other in solving simple numerical puzzles known as "sudokus". The important feature of our experimental design is that one competitor in each pair faces a naturally induced disadvantage in the respective task.[4] This difference is due to the fact that students in one school ("experienced") are taught how to solve sudokus as part of their regular math classes, while students in the other school ("non-experienced") are not. The schools are very similar in all other relevant aspects: both are private, located in similar neighborhoods, fully bilingual and have good records in national math and science competitions. Therefore, the difference in experience can be regarded as an exogenous source of disadvantage imposed on two similar groups, since practice in solving sudokus in the experienced school was not yet common at the time most parents made their schooling decision.[5] We implement two different types of AA policies, lump sum and proportional bonuses, to compensate for this disadvantage in capacities to compete.[6] Hence, we analyze affirmative action policies in a situation where a real and systematic disadvantage between competing individuals is present. Admittedly, the described set up is somehow artificial since AA is usually implemented in situations where the differences between favored and non-favored groups are more ambiguous and complex, such as those derived from racial or ethnic differences, stereotype threats or

[4] Coate and Loury (1993) show theoretically how discrimination may arise in two symmetric groups as a self-fulfilling prophecy. We take such asymmetry as given.

[5] We do not compare behavior *between schools* but *between treatments in the same school*. Hence, similarity between schools is not a necessary condition for the validity of our analysis. However, the justification of affirmative action is based on the premise that the only relevant difference between schools is the ex-ante experience in solving sudokus.

[6] Calsamiglia (2009) shows that an appropriately designed AA policy should equalize rewards to effort whenever the set-up affects one of many factors determining individual final welfare. In this particular environment rewards to effort are equalized with proportional AA.

different propensities to compete across different genders.[7] Instead, our design allows us to identify the incentive effects of affirmative action resulting from the reduction of an existing asymmetry in actual capacities to compete, thereby abstracting from the intricate nature of discrimination in actual applications. Our study suggests that reducing identified asymmetries to compete may improve performance.

Our tournament is in line with Schotter and Weigelt (1992), who studied the incentive effects of AA in a laboratory experiment where the choice of a monetary cost symbolizes the choice of exerted effort in a tournament. In their experiment, subjects' exogenous disadvantage was induced by assigning different cost parameters for which individuals were later compensated for through affirmative action, implemented as lump-sum bonuses of different sizes. Their results indicate that AA can either boost or worsen performance depending on the relative size of the cost disadvantage with respect to the implemented compensation. Our experiment provides evidence showing that AA can enhance effort exertion and therefore performance of participants in a real effort framework where the disadvantage is due to a de-facto asymmetry in capacities to compete.

Subjects in our experiment were school children and the tournament was conducted in their respective schools. Working with children in their natural environment offers a clean setup to analyze students' behavior when exposed to exogenously different experiences imposed by their schools. Using children as subjects has additional advantages: they react rationally, spontaneously and in line with economic theory; their performance is not affected by them questioning the underlying motivation of the experimentalist; and it is relatively easy to provide them with incentives, see Harbaugh et al. (2001) and Harbaugh and Krause (2000). Finally, studying how children react to AA is important since some social asymmetries may be ideally resolved at early ages, before they are exacerbated.

The implementation of AA is usually accompanied by an intense public debate on whether such policies satisfy certain fairness criteria and on the possible inefficiencies they may create.

[7] Some recent experimental studies analyze the consequences of different types of naturally occurring differences among social groups in competitive situations. For instance, Hoff and Pandey (2006) study stereotype threats among members of different social castes in India. Also, Gneezy et al. (2003), and Niederle and Vesterlund (2007), study the different propensity to compete of women and men in mixed-gender tournaments. A recent study that analyzes AA based on the mentioned gender difference is Niederle, Segal and Vesterlund (2010), where their focus is on the effects of gender quotas on women's participation decisions and performance in tournaments.

This latter concern derives from the underlying assumption that the discriminated group is often less qualified. Under such assumption, favoring members of the disadvantaged group in the selection process may not only result in a more balanced representation of disadvantaged individuals but also in a lower overall performance in the selected group. An important strand of the literature has investigated the effects of AA policies in these two dimensions. Holzer and Neumark (2000) survey this literature and argue that, while AA increases representation, results with respect to the performance of the selected group are mixed. More recently, McCrary (2007) shows that AA increases the representation of blacks in the police force, but that their performance does not decrease in a significant way. Bertrand, Hanna and Mullainathan (forthcoming) analyze the effects of AA in college both in terms of the increased representation and their success in college and in the labor market. Miller and Segal (2008) study the effect of AA on the pool of hired law enforcement officers.

These studies take performance by individuals as given and focus their analysis on the performance distribution of the newly selected pool of individuals. But more often than not, qualification and therefore performance in a position is a result not only of ex-ante ability, but also of an investment made by the individual. Deciding how much to invest will be determined by the incentives induced by the selection policies, that is, by the chances of being selected, and in particular by AA policies. Many selection processes, such as college admission or the labor market, induce effort choices, *prior* to the selection per se, which crucially affect the observed performance distribution of the selected group. If skill on a specific task is a result of initial ability, but also of effort exerted to learn and improve performance, then the final performance distribution will depend on both, the initial distribution of abilities of competing agents, and the incentives provided by the selection rules. For instance, if we want to select individuals with high SAT scores, then the hours they have invested in studying will clearly affect their final performance and therefore the performance of the admitted group of individuals. Understanding the incentives provided by the selection rules is crucial for the success of the process.[8]

[8] The implementation of AA might also alter the participation decisions of competitors which additionally affects effort incentives for participating agents. For instance, Marion (2009) and Krasnokutskaya and Seim (2010) analyze how bid preference programs for minority-run businesses in public procurement auctions can affect the set of participants and the final cost paid for a given project. Our experimental design is based on compulsory pair-wise tournaments which implies that subject participation is not an issue as both competitors will be active participants.

There also exist a number of theoretical studies that analyze the effects of balancing the ex-ante heterogeneity of competing individuals. For instance, Lazear and Rosen (1981) show that a handicapping system induces efficient competition in a rank-order tournament between weak and strong players. Che and Gale (1998) analyze bid caps for strong bidders in an all-pay auction framework. Also Myerson (1981) shows that an optimal, i.e., revenue maximizing auction between asymmetric bidders implies favoring weak bidders. Theoretical models that explicitly address the incentive effects of affirmative action policies can be found in Schotter and Weigelt (1992), Fu (2006), Franke (2008), Balart (2009) and Hickman (2010).[9] In those papers affirmative action is modeled as a bias in favor of ex-ante disadvantaged players in an all-pay auction or contest set-up. The conclusion that can be drawn from most of these papers is that reducing the asymmetry in competitive advantage tends to enhance individual performance. Whether the implementation of affirmative action policies in real applications is suitable to incentivate individuals remains an open empirical question.

In our empirical analysis of the experimental data we find some evidence for the incentivizing effects off AA. Based on two benchmark treatments we first confirm the crucial assumption of our set up, that is, that the asymmetry in experience is reflected in subjects' performance. In fact, subjects from the school where solving sudokus was part of regular class activities had an advantage in the competition: they solved significantly more sudokus than non-experienced subjects in a benchmark treatment where this difference in experience was not revealed to the tournament participants. Moreover, making this asymmetry in experience publicly known did not lead to significant changes in performance of neither, experienced and non-experienced individuals.

Regarding the incentive effects of AA we find that the performance of non-experienced subjects was enhanced independently of their ability. However, for experienced individuals performance effects differed with ability: for subjects with relatively low or average ability, performance was higher in the AA treatments while for those with highest ability performance was slightly lower compared to the benchmark treatment without AA. Those effects were relatively robust with respect to the two different types of AA policies we implemented in our

[9] See also the general discussion in Fryer and Loury (2005).

tournament, i.e., lump-sum and proportional compensation, but were more pronounced for those policies where the bias was relatively high.

Moreover, AA policies balanced the tournament on average, since around half of non-experienced subjects in the AA treatments won their respective tournaments in all potential matches. Also, the average performance of all potential tournament winners selected through AA was only moderately lower than the average performance of the winners who would have been selected without it. Hence, the negative selection effect of selecting a higher proportion of non-experienced individuals as tournament winners was partially compensated by increased levels of effort performed under AA. We also find that AA positively increased the confidence in winning of non-experienced subjects, while that of experienced subjects was unaffected.

Finally we briefly address the perception of the affected subjects with respect to the different AA policies. While all subjects regarded the initial asymmetry in experience as unfair, fairness perceptions varied with treatments. For instance, proportional AA, whose size depends on individual performance, was perceived as fairer than lump-sum AA, although the actual compensation received in the proportional treatments was on average higher than in the lump-sum treatments.

The rest of the paper is organized as follows. Experimental design and procedures are explained in Section 2. Section 3 presents the results. Section 4 sums up our conclusions. The Appendix contains an English translation of the instructions used in the experiment.

2. **Experimental Design and Procedures**

We conducted pair-wise tournaments among 336 school children, aged 10-13, from two similar non-religious, bilingual private schools located in the similar neighborhoods in Barcelona. Students at both schools have a systematic difference in experience in a specific real-effort task consisting in solving simple “sudokus”. This ex-ante difference in experience is due to the fact that during regular math classes, students in the “experienced” school (E) are trained in solving sudokus (and in fact have to solve sudokus as part of their regular homework) while students at the "non-experienced" school (NE) are not.[10]

Sudoku is a logic-based number-placement puzzle. The objective is to fill a 9x9 grid so that each column, each row and each of nine 3x3 boxes contains one-digit numbers from 1 to 9 only once. The puzzle setter provides a partially completed grid. We use a simplified 4x4 grid version in order to obtain sufficient variation in performance. We chose this task because the rules are simple, yet it requires substantive logical reasoning and concentration by the subjects. Additionally, performance is easy to measure and, crucially, depends on effort. Most importantly, both effort and ability play a role, so that non-experienced subjects still have a chance of winning, independently of whether they are favored by an affirmative action policy or not.[11] Figure 1 below shows one of the sudokus used in the experiment (a) and its solution (b).

	4		2
		3	
1			

(a) Unsolved Sudoku

3	4	1	2
2	1	3	4
1	2	4	3
4	3	2	1

(b) Correctly Solved Sudoku

Figure 1: An example of the real-effort task (sudoku).

Each student from E was randomly and anonymously matched with a student from NE in his or her same school year (4th or 6th grade). Each pair competed in a tournament which lasted 30 minutes.[12] Subjects had to correctly solve as many sudokus as possible in order to beat their matched rival. All subjects were handed the same answer sheet containing 96 sudokus randomly

[10] An ex-post experimental questionnaire showed that a small fraction of students from both schools were familiar with sudokus due to prior experience outside school. We control for this ex-ante experience in our analysis by using a proxy for ability. The task was defined as "filling in a grid" and the word "sudoku" was never mentioned.

[11] In fact, the percentage of NE winners in their respective tournament was at least 13.3% (for experimental treatment "K" and 4th year students, where no affirmative action was implemented).

[12] We chose pair-wise tournaments instead of multiple-prize tournaments with N players because the schools did not allow us to establish intra-school competition. Additionally, pair-wise tournaments allowed us to control the amount of information that each subject had on its rival's ability.

generated with the same level of difficulty by a computer program.[13] Each pair of subjects was competing for a 7€ voucher from a bookshop located in Barcelona.[14] In each pair, the student who had correctly solved more sudokus than the respective opponent during a 30 minute period won the voucher. In the case of ties, the winner was determined randomly.

Once we verified that experience provided an advantage for subjects in the experienced school, our objective was to study: 1) the effect of providing information on competitors' previous experience with the task; and 2) the effect of implementing affirmative action policies on subjects' performance and as a result, on the output generated by subjects selected as tournament winners. Thus, we randomly assigned similar numbers of subjects from each school to each of six treatments. In treatment NK no subject was informed about whether subjects from the other school were experienced or not in solving sudokus. In treatment K students at the NE school were told that students in the E school had previous experience in solving sudokus. Similarly, students in the E school were told that students at the NE school were not taught how to solve sudokus. In the remaining four treatments all subjects were informed about the existence of a difference in experience across schools and about the particular affirmative action policy applied to compensate NE subjects. In treatments LH (Lump-sum High) and LL (Lump-sum Low), all subjects were informed that NE subjects were given a predetermined number of solved sudokus ex-ante: 20 in LH and 8 in LL. In treatments PH (Proportional High) and PL (Proportional Low), all subjects knew that NE subjects were given a number of solved sudokus proportional to the number of sudokus they correctly solved, one for every correctly solved sudoku in the case of PH, and one for every two correctly solved sudokus in the case of PL. Table 1 summarizes our treatment design.

Table 1: Description of Treatments

Not Know	NK	Subjects unaware of others' experience
Know	K	Subjects aware of others' experience
Lump-sum High	LH	Subjects aware of experience and NE subjects receive a bonus of 20 correct sudokus bonus
Lump-sum Low	LL	Subjects aware of experience and NE subjects receive a bonus of 8 correct sudokus bonus
Proportional High	PH	Subjects aware of experience and NE subjects receive 1 correct sudoku bonus for every 1 correct

[13] The software used was "SuDoku Pro" by Dualogy Systems. The proportion of mistakes across all solved sudokus was not statistically significant. No subject was able to complete all 96 sudokus provided.

[14] Subjects were explicitly told that the voucher was redeemable for "books, collector's cards, toys, music or comics". Experiments took place at approximately the time the final *Harry Potter* book, priced at 20 Euro, was published in Spain.

Proportional Low	PL	Subjects aware of experience and NE subjects receive 1 correct sudoku bonus for every 2 correct

Based on this treatment design we can address our research questions as follows. The comparison of performance in treatment NK between schools E and NE allows us to see whether students from the E school were in fact advantaged by their ex-ante experience. Comparisons of performance across treatments NK and K within a school allow us to study the effects of making the information on different levels of ex-ante experience salient. Finally, through the comparison of within-school performance between treatments K and the different treatments in which AA is implemented (LL, LH, PL, PH) we can assess the isolated effect of affirmative action, abstracting from the information effect. In parts of our analysis, we pool the data from all treatments where affirmative action is implemented and refer to them generically as the "AA" treatment.

The sizes of the affirmative action policies were determined using results from pilot experiments with children of similar age in other schools where solving sudokus was never part of regular course activities. In such pilots we adopted a 15 minute time period without any AA policies and asked for private ex-ante experience with sudokus. The difference in the number of correctly solved sudokus between privately experienced and non-experienced subjects was roughly 4 sudokus. In the main experiment we doubled the time span to 30 min and set the LL bonus such that this difference in experience would have roughly been leveled out. The LH bonus of 20 sudokus was close to the average number of correctly solved sudokus by our pilot non-experienced subjects. The PL and PH bonuses were designed to be close to the respective values for the lump-sum treatments for an average non-experienced participant.

Prior to conducting the experiments, we repeatedly met with faculty from both schools in order to guarantee their collaboration and pedagogical interest in the project. During these meetings we obtained information on subjects' gender, birth date, teaching group and school grades. We later assigned subjects to treatments in such a way that the groups were balanced in accordance with these pre-specified characteristics. Table 2 below shows descriptive statistics of subjects assigned to each treatment at each school.[15] Small variations across treatments were

[15] Average Grade is calculated using grades in all topics in the preceding term and it is slightly higher at the NE school than at the E school (3.55 vs. 3.34, significant at the 1% level). This difference may be due to different grading systems across schools.

mainly due to absent students and latecomers.[16] Participation was mandatory, which helped to avoid selection biases and simplified matters for the school. None of the subjects manifested opposition to participating.

Table 2: Description of the Subject Pool

N = 336	Experienced						Non-experienced					
	NK	K	LH	LL	PH	PL	NK	K	LH	LL	PH	PL
% Female	41	43	48	53	50	48	46	46	48	39	48	47
% 6th Year	48	43	45	43	46	48	50	46	59	48	48	59
Average Grade (1=Worst,5=Best)	3.32	3 .31	3.47	3.46	3.14	3.35	3.44	3.44	3.65	3.65	3.54	3.57
Number of subjects	29	30	31	30	28	31	24	24	27	23	27	32

Experiments were carried out on two separate but close dates in 2008. In each school experimental sessions took place at different times of the day for 4th and 6th graders for practical reasons.[17] Subjects were conducted to separate classrooms according to our predefined assignment. While students waited for the experimentalist, teachers conducted a specific and identical school activity (writing an essay) in order to keep the subjects calmed and equally uninformed about the experiment. The same experimentalist arrived at each of the classrooms at twenty-minute intervals and then sessions started.[18] Teachers were not present during the experimental sessions, in order to minimize their influence.

The experimental sessions lasted one hour. First, the experimentalist read out general instructions on how to solve sudokus (see "Pre-instructions" in the Appendix). Then, subjects had a five-minute practice round to solve sudokus with no incentives being offered and no mention of competition.[19] After this period, the experimentalist solved one of the practice

[16] If there were unequal number of participants in one treatment in the two schools we used the results of randomly determined and already matched subjects to determine the tournament winner for the remaining unmatched subjects. However, each subject could only win once even it was randomly determined. This is also the reason why the instructions only specify the consequences for the respective subject and not for his/her direct rival.

[17] We are unaware of cross-contamination between schools or between subjects from different school years at the same school. The timing of the experiments was carefully designed so as to avoid these problems.

[18] This was the reason why different treatments were carried out at different time-intervals. Since the experiment deals with effort motivation and children may be easily influenced, it was crucial to have the same experimentalist conducting the sessions. The experimentalist rehearsed repeating exactly the same cues across sessions.

[19] We will use the number of correctly solved sudokus in this practice round as a proxy variable for ability ("Pretest"). There is a strong positive correlation between this variable and tournament performance in all treatments (always above 0.7).

sudokus in front of the students. Once questions were clarified, instructions for each of the treatments were read aloud. The instructions made it clear that each student was competing against an anonymous student from another comparable school and that students at the other school were systematically experienced (or not) in solving sudokus (for treatment NK this information was omitted). The difference in ex-ante experience was explicitly mentioned to justify the implementation of the affirmative action bias in favour of the non-experienced group in the AA treatments (see the Appendix for the instructions). Tournament rules were explained giving numerical examples (specific to each treatment) for all potential outcomes of the tournament, i.e., losing, winning, and tying. Moreover, aggregate information with respect to the number of sudokus (i.e., mean, minimum and maximum) that had been correctly solved by a comparable subject pool was provided.[20] This information, identical for all subjects, was based on the results of our pilot experiments. The experimentalist also held up a 7€ voucher to increase the credibility of the prize offered to tournament winners. After that, subjects had thirty minutes to solve the sudokus in two separate handouts. After the first fifteen minutes, subjects were instructed to start working on the second handout, so that we could measure whether there were intra-session learning effects or whether these were over-ruled by fatigue.[21] Subjects were explicitly told that they could stop solving sudokus at any time and start any other activity, such as drawing, under the condition that they remained quiet and did not bother others. In fact, very few subjects (less than two per treatment) quitted.

After the thirty minutes had passed, the handouts were collected and a questionnaire about previous experience in solving sudokus, self-confidence and the perceived fairness of the implemented affirmative action policies was distributed. Once the questionnaires had been filled in, subjects continued with their regular classes. The experimentalists then randomly matched participants from both schools, determined the winners and deposited the vouchers at the schools, so that they could be distributed by school faculty.

3. Results

[20] The empirical analysis revealed that the specific numbers mentioned in this information did not lead to an anchoring effect for the subjects.

[21] We did not find substantive differences in performance between the two parts of the test, indicating that the effects of learning and fatigue possibly cancel out. Experienced subjects completed on average one more sudoku in the second part than in the first part (significant at the 1% level). Non-experienced subjects did not solve a significantly different number of sudokus in the two parts.

3.1 Descriptive Statistics

We start by taking a descriptive look at the data. Table 3 reports the average number of correct sudokus by treatment and school year (4th or 6th grade) in each of the schools (E and NE), as well as standard deviations (in parentheses). There is high heterogeneity in performance in all treatments and schools and thus, standard deviations are large. Table 3 provides a first indication that experienced subjects (E) solve, on average, more sudokus than non-experienced subjects (NE), a key hypothesis justifying our experimental design. Aggregating over all treatments, experienced subjects solved 36.06 sudokus, while non-experienced subjects solved 23.31 sudokus.[22] Using the number of sudokus solved in the five minute practice round as a measure of individual ability in solving sudokus, we find that experienced subjects of low ability, those who solve three sudokus in the practice round, solve a similar number of sudokus (25.61) in the tournament as the average of non-experienced subjects.[23]

Table 3 also shows that age affected performance. The average performance of 4th grade experienced subjects in all treatments is similar to that of 6th grade non-experienced subjects.

Table 3: Average Number (and Standard Deviations) of Correct Sudokus by Treatment and School Year

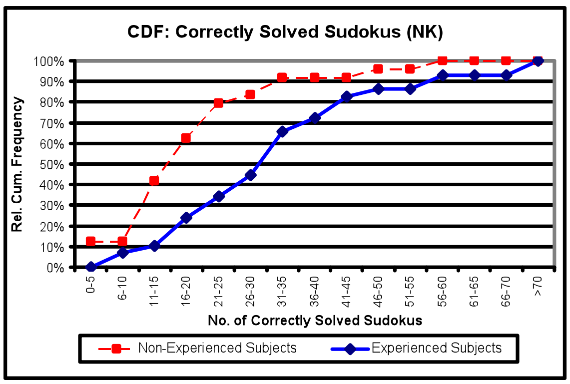
	4th Grade		6th Grade		Overall	
	E	NE	E	NE	E	NE
NK	**28**	**16.08**	**38.93**	**24.67**	**33.27**	**20.38**
	(15.43)	(8.01)	(16.10)	(15.44)	(16.44)	(12.80)
K	**29.88**	**17.69**	**43**	**29.09**	**35.57**	**22.92**
	(12.47)	(10.74)	(17.98)	(13.43)	(16.22)	(13.13)
AA	**29.38**	**19.26**	**45.67**	**28.08**	**36.85**	**24.04**
	(13.78)	(9.48)	(12.04)	(12.12)	(15.31)	(11.80)
LH	27.59	23.36	44.86	29.50	35.39	27
	(13.13)	(9.19)	(11.51)	(14.43)	(15.02)	(12.73)
LL	27.59	19.42	51.54	26	37.97	22.57
	(12.26)	(11.79)	(11.44)	(9.01)	(16.82)	(10.85)
PH	29.67	17.93	46.92	26.54	37.68	22.07
	(12.38)	(9.05)	(11.09)	(11.16)	(14.52)	(10.84)
PL	32.94	17.07	40.27	29.16	36.48	24.25

[22] Moreover, non-experienced subjects solved on average less sudokus than experienced subjects in each separated treatment, a clear indication that the disadvantage was real and not simply a self-fulfilling expectation.

[23] We will use the number of correctly solved sudokus in the trial round as a proxy for ability in the respective task. During the experimental sessions there was no indication that subjects did not try hard during the five minute trial. Moreover, the coefficient of this measure of ability ("pretest") in the regression analysis presented later is always highly significant for sudoku performance.

	(17.36)	(7.60)	(12.38)	(12.83)	(15.37)	(12.42)

Figure 2 below shows the cumulative distribution function (CDF) of the number of correct sudokus solved by students in the E and NE school in the two benchmark treatments NK and K. Note that the distributions have a large spread and range from 0 correctly solved sudokus to more than 70. Stochastic dominance of the CDFs for the E school clearly shows that the lack of experience in solving sudokus is in fact a disadvantage for the NE subjects. Mann-Whitney tests comparing the inter-school number of correct sudokus in both of these treatments show significant differences at the 1% level (p-values of 0.002 for NK and of 0.004 for K).



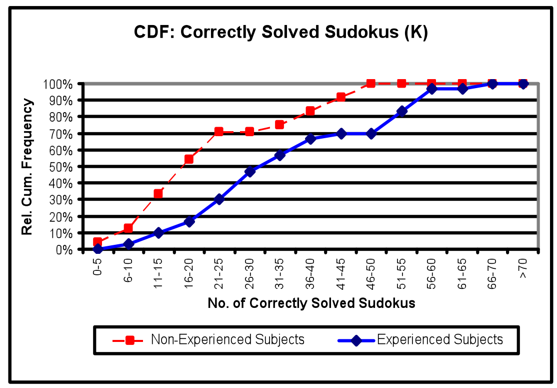


Figure 2: CDFs of the number of correct sudokus by E and NE in the NK and K treatments.

Intra-school comparisons across treatments are less clear-cut. Table 3 reports the number of correctly solved sudokus for each of the four separate AA treatments (LH, LL, PH and PL), as well as for the pooled treatment (AA).[24] Although standard deviations are very large, the averages give a first indication that performance may be enhanced when providing information (K vs. NK treatments) and that affirmative action policies also enhanced performance (AA vs. K treatments, with the only negative comparisons being for E subjects in 4th grade and NE subjects in 6th grade). Figure 3 depicts the CDFs for the number of correct sudokus in the two benchmark treatments NK and K, as well as the pooled AA treatment separately for E and NE subjects. Visually, the CDF for the K treatment "almost stochastically dominates" the CDF for the NK treatment in both graphs, suggesting that the provision of information about the disadvantage for one group of subjects did not decrease performance. Similarly, the CDF for the AA treatment also lies mostly below the CDF for the K treatment in both schools which implies that subjects

[24] Kruskall-Wallis tests comparing the four AA treatments in each school do not show significant differences among treatments. Thus, we will use the pooled AA treatment in this section.

faced with AA policies do not decrease their performance. Comparing the distributions of all treatments based on pair-wise Mann-Whitney tests does not generally result in significant differences at the standard levels, apart from the comparison of NK with LL for 6th year experienced subjects (average of 38.93 correct sudokus in NK and average of 51.54 in LL, p-value of 0.01).

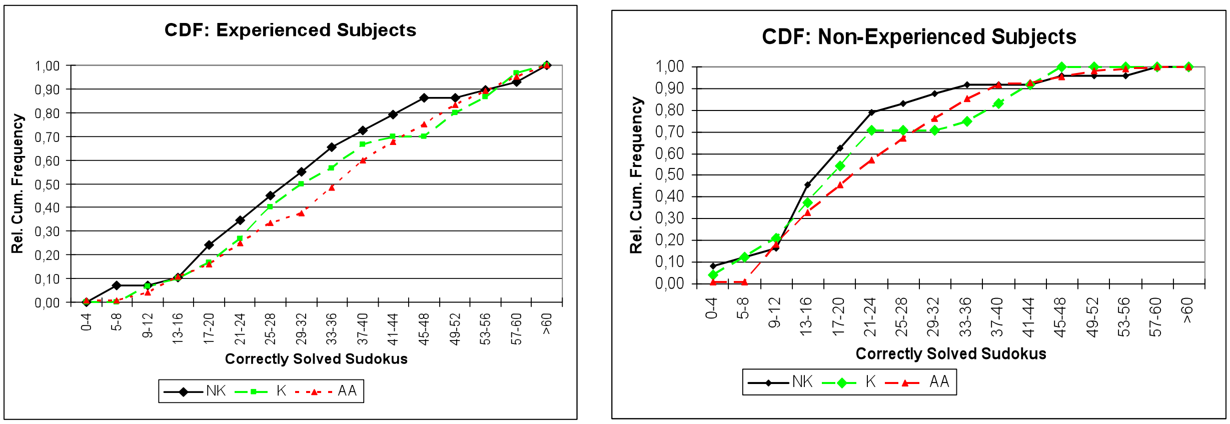


Figure 3: CDFs of the number of correct sudokus by school in the NK, K and AA treatments.

The fact that statistical tests are mostly non-significant can be attributed to large heterogeneity among subjects, which cannot be controlled for by using non-parametric methods. Therefore, in the next section we rely on regression analysis as it allows us to incorporate substantial additional information on the subject level to control for unobserved individual heterogeneity.

3.2 The Effects of Information on Performance

In this subsection we analyze whether providing the subjects with information about the previous experience of their respective rivals affected performance. Therefore we compare the number of correct sudokus in the K and NK treatments. Table 4 below provides results from OLS regressions with robust errors restricted to the two benchmark treatments NK and K, separately for experienced and non-experienced subjects.[25] In both regressions the dependent variable is the number of correct sudokus while explanatory variables are a dummy variable for the K treatment

[25] The range and the variance of the dependent variable are sufficiently large such that OLS regressions should be appropriate. Moreover, our specification is robust with respect to the use of Poisson regressions: results are similar with respect to sign and significance levels of the coefficients (the size of coefficients is not directly comparable between these two approaches as marginal effects in Poisson models are not constant)

("K"), the number of correct sudokus in the trial phase ("Pretest") [26], the average grade obtained by each student in the preceding term ("Grade"), a dummy variable to indicate 4th or 6th grade ("Year") and a dummy variable for gender ("Gender").

Table 4: Correct Sudokus and Information		
	Experienced	Non-Experienced
	REG (1) Dep. Var: # Correct Sudokus	REG (2) Dep. Var: # Correct Sudokus
Constant	-9.68 (5.87)	1.60 (6.20)
K	-2.57 (2.65)	-0.06 (2.33)
Pretest (0=Min, 6=Max in E) (0=Min, 12=Max in NE)	7.22 (0.86)***	3.14 (0.95)***
Grade (1=Worst,5=Best)	3.11 (1.43)**	1.33 (1.89)
Year (0=4th,1=6th)	9.22 (2.49)***	6.62 (3.12)**
Gender (0=Male,1=Female)	0.69 (2.40)	1.70 (2.51)
# Observations	59	47
Adj. R^2	0.70	0.58

Notes: * denotes significance at the 10% level, ** at the 5%, and *** at the 1% level. Robust standard errors are in parentheses.

The results of both regressions suggest that there is no effect of providing information on performance.[27] Coefficients for the K treatment are negative in both cases but not significantly different from zero (and actually very close to zero for non-experienced subjects). Most of the other controls, apart from gender, are important in explaining performance and have the expected sign. We summarize our results as follows:

[26] In the trial phase experienced subjects were provided with only 6 sudokus. Therefore, our measure of ability for experienced subjects is cut off at 6, since it includes individuals who would possibly have solved more than 6 sudokus. We thus expect our estimated parameters to be smaller and less significant (due to higher variance). 40% of experienced subjects solved all six sudokus correctly. For the subsequently run experimental sessions with non-experienced subjects, we extended the number of trial sudokus to 12. By truncating these data artificially in the same manner as for experienced subjects, we were able to verify the conjecture that with fewer sudokus in the trials results become slightly less significant and weaker in absolute size without altering the qualitative results.

[27] This observation has been documented in other studies based on recent real-effort tournaments. A recent example is Freeman and Gelber (2009), where performance is not substantially altered in their single prize tournament when competitor's past performance is revealed.

Result 1: *Knowledge of the existence of an asymmetry in experience did not change performance by experienced or non-experienced subjects.*

3.3 The Effects of Affirmative Action on Performance

In this subsection the incentive effects resulting from the implementation of AA policies are addressed based on separate regressions for the two subject pools of experienced and non-experienced subjects. Table 5 presents the results from OLS regressions with robust and clustered errors for experienced subjects while Table 6 contains the results for non-experienced subjects. Our baseline treatment in both cases is benchmark treatment K in which subjects are aware of the existing disadvantage.[28] Explanatory variables are the four AA treatment dummies, which are additionally pooled in separate regressions, in addition to the other controls already used in regressions (1) and (2) such as a proxy for unobserved ability ("Pretest"), school grades ("Grade"), and a dummy for being in sixth grade instead of fourth grade ("Year"). All these controls have positive and significant coefficients,[29] with the exception of the "Gender" variable.[30]

Subjects might react differently to implemented AA policies depending on their respective ability. Therefore, we include an additional interaction term ("AA*Pretest") to capture these effects in REG (3) and (4) for experienced, as well as in REG (7) and (8) for non-experienced subjects. The inclusion of the interaction term implies that a representative subject of the base group in treatment K has low ability (zero correct sudokus in the five minute trials).

Results for experienced subjects in Table (5), REG (3), seem to imply that AA does not have a significant effect on the number of correct sudokus. However, REG (5) reveals that subjects of different abilities react differently to the implementation of AA. While the coefficient for AA is

---

[28] All results are maintained if treatment K and NK are used as baseline treatment and an additional dummy variable for treatment NK is added.

[29] "Grade" is not statistically significant for NE subjects.

[30] There exists an important literature analyzing how male and female individuals react differently to competition, see Gneezy et al. (2003), Gneezy and Rustichini (2004), and Niederle and Vesterlund (2007). Subjects in our experiment did not know the gender of their respective rival, which might be an explanation of the insignificance of the gender variable in our experiment. A related study that addresses this issue in a gender context is Niederle, Segal and Vesterlund (2008) where they find that gender quotas in tournaments increase the probability that female subjects decide to enter into competitive tournaments.

large, positive, and significant at the 5% level in this specification, the respective interaction variable "AA*Pretest" is negative and also significant. These results imply that experienced subjects in the base group (with low ability) statistically solve 9.14 more sudokus when they compete with subjects favored by affirmative action policies. However, the negative and significant coefficient of the interaction term indicates that the higher the ability of experienced subjects (measured by "Pretest"), the lower the increase in performance under AA.[31] Notice that these results imply that the performance of experienced subjects of highest ability under AA is slightly lower than the performance of experienced subjects in the baseline treatment.

Table 5: Correct Sudokus and Affirmative Action among Experienced Subjects

	REG (3) Dep. Var: # Correct Sudokus	REG (4) Dep. Var: # Correct Sudokus	REG (5) Dep. Var: # Correct Sudokus	REG (6) Dep. Var: # Correct Sudokus
Constant	-7.61 (3.06)**	-8.33 (3.05)**	-14.22 (3.74)***	-14.47 (3.84)***
AA	0.88 (1.59)	**-**	**9.14 (3.77)****	**-**
AA*Pretest	**-**	**-**	**-1.80 (0.78)****	**-**
LH	-	0.95 (2.01)	-	11.42 (6.10)*
LL	-	-0.01 (2.20)	-	-2.30 (4.63)
PH	-	3.37 (1.39)**	-	13.37 (4.25)***
PL	-	-0.60 (2.69)	-	4.60 (6.12)
LH*Pretest	-	-	-	-2.35 (1.08)**
LL*Pretest	-	-	-	0.39 (0.92)
PH*Pretest	-	-	-	-2.20 (0.86)**
PL*Pretest	-	-	-	-1.13 (1.49)
Pretest (0=Min, 6=Max)	5.69 (0.49)***	5.76 (.047)***	7.12 (0.54)***	7.08 (0.56)***
Grade (1=Worst,5=Best)	3.69 (0.82)***	3.81 (0.82)***	3.65 (0.80)***	3.83 (0.76)***
Year (0=4th,1=6th)	10.21 (1.58)***	10.16 (1.56)***	10.63 (1.63)***	10.21 (1.57)***
Gender	0.84	0.86	0.82	0.77

[31] The statistical effect of AA on subjects with higher ability can therefore be calculated as "AA"+"AA*Pretest".

(0=Male,1=Female)	(1.66)	(1.66)	(1.67)	(1.65)
# Observations	150	150	150	150
Adj. R^2	0.65	0.66	0.66	0.67

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** at the 1% level.

Robust standard errors, clustered by treatment and class are in parentheses.

REG (4) and (6) allow us to analyze whether the established results for the pooled AA treatment are driven by specific AA policies. While REG (4) suggests that the PH treatment is the only significant AA treatment (at the 5% level), REG (6) indicates that the AA treatments with large size, i.e., treatments LH and PH, have significant positive effects (at the 10% level for LH and 1% level for PH) when accounting for the interaction of our proxy for ability and the different AA treatments.[32] Moreover, and in line with the results obtained from the pooled AA treatment in REG (5), the coefficients of the interaction term with respect to the intensive AA treatments ("LH*Pretest" and "PH*Pretest") are negative and significant at the 5% level. This suggests that the more intensive LH and PH treatments are the main contributors to the derived incentive effects. [33] We thus conclude:

Result 2: *Affirmative Action policies enhanced the performance of experienced subjects with the exception of subjects of highest ability. Most of the effect can be attributed to treatments where the compensation was high, i.e., PH and LH.*

We now focus our analysis on the group of non-experienced subjects. Results from the same specification used for the experienced subject pool are presented in Table 6. REG (7) reveals that pooled AA has a positive and significant effect (at the 5% level) on performance, i.e., non-experienced subjects solve on average 3.98 more sudokus when they are favored by an affirmation action policy. REG (8) suggests that the incentive enhancing effects of affirmative action cannot be attributed to a specific type of policy because coefficients for all AA treatments (LH, LL, PH and PL) are similar in size and significance levels (5% in the case of LH, 10% in the others). Hence, there seems to be a performance enhancing effect of affirmative action policies on non-experienced subjects which is independent of its form or size. The lack of

[32] Results are confirmed if the LH and PH treatments are pooled under one "high" variable and the LL and PL treatment under a "low" variable.
[33] Using the total number of solved sudokus as dependent variable (instead of the number of correctly solved sudokus) leads to similar results in all four regressions that are available upon request.

sensitivity to the specific type and size may be the result of non-experienced subjects' lack of familiarity with the task, which may reduce their capacity to assess the relative size of compensations. The interaction term between AA and Pretest is included in REG (9) for the pooled AA treatments and in REG (10) for each AA treatment separately. In contrast to the group of experienced subjects there is no indication of different reaction by low or high ability subjects to AA because the interaction terms remain insignificant (apart from the positive and significant effect of LH*Pretest at the 5% level). This might be attributed to the fact that non-experienced subjects have less ground to assess their relative ability, since the task is new to them.[34]

Table 6: Correct Sudokus and Affirmative Action among Non-Experienced

	REG (7) Dep. Var: # Correct Sudokus	REG (8) Dep. Var: # Correct Sudokus	REG (9) Dep. Var: # Correct Sudokus	REG (10) Dep. Var: # Correct Sudokus
Constant	3.01 (3.59)	3.16 (3.62)	8.69 (4.24)**	8.74 (4.30)**
AA	**3.98** **(1.66)****	-	-3.10 (3.33)	-
AA*Pretest	**-**	-	1.84 (1.19)	-
LH	-	4.71 (2.03)**	-	-5.55 (3.83)
LL	-	3.79 (2.01)*	-	-3.54 (4.09)
PH	-	3.64 (1.84)*	-	-1.89 (3.52)
PL	-	3.76 (2.18)*	-	-2.10 (4.16)
LH*Pretest	-	-	-	2.55 (1.24)**
LL*Pretest	-	-	-	1.97 (1.54)
PH*Pretest	-	-	-	1.40 (1.37)
PL*Pretest	-	-	-	1.48 (1.32)
Pretest (0=Min, 12=Max)	4.25 (0.60)***	4.23 (0.62)***	3.04 (1.19)**	3.03 (1.22)**
Grade (1=Worst,5=Best)	0.03 (0.87)	0.03 (0.89)	-0.13 (0.82)	-0.12 (0.82)
Year (0=4th,1=6th)	4.45 (1.61)***	4.42 (1.65)**	4.05 (1.36)***	4.24 (1.45)***

[34] The limited sample size for each treatment in combination with large heterogeneity in performance of subjects with different abilities might be an alternative explanation for the non-significant results in REG (9) and (10). Using the total number of solved sudokus as dependent variable in REG (9) and (10) leads to similar sized coefficients which are now mostly significant in these specifications. These results are available on request.

Gender (0=Male,1=Female)	0.46 (1.35)	0.43 (1.37)	0.66 (1.39)	0.31 (1.53)
# Observations	132[a]	132[a]	132[a]	132[a]
Adj. R^2	0.65	0.65	0.67	0.67

(a): For one non-experienced subject "Grade" was not available. Another subject arrived late and did not participate in the practice rounds. Such observations are omitted.

Nevertheless, the significant and positive effects of AA in REG (7) and (8) allow us to derive the following result:

Result 3: *Affirmative Action policies enhanced the performance of non-experienced subjects independently of the two different sizes and types of the implemented policy.*

3.4 The Effects of Affirmative Action on the Selection of Tournament Winners

An important concern raised in the context of affirmative action policies is that the selected pool of candidates may be of lower ability because of the higher proportion of selected disadvantaged individuals who may perform poorly. There are two different approaches to answer this question, crucially depending on the objective of these policies. First, the objective may be to select individuals according to unobserved ability, which is assumed to be equally distributed among advantaged and disadvantaged individuals. If this assumption holds and ability is positively correlated with performance then selecting a similar proportion of the best performing individuals in both groups should lead to a pool of selected individuals that consists of the highest ability individuals overall. In our benchmark treatments NK and K, where no affirmative action is implemented, the proportion of non-experienced subjects among all possible tournament winners[35] amounts to only 24%. However, the percentage of non-experienced subjects among possible tournament winners is substantially higher under the AA treatments: it increases to 51%, reflecting that the objective of selecting a more balanced sample of tournament winners, consisting of those subjects who irrespective of their training have highest inherent ability, is met on average. This result shows that the implemented AA policies on average leveled the playing field.

Second, if the objective of the tournament is to select the highest performing individuals, then the average performance of tournament winners may be lower under AA, since a higher

[35] To find the possible tournament winners we computed the mean among all possible matches within each treatment. Note that the particular match used to reward subjects in our experiment was just one random realization of this process.

number of disadvantaged subjects are selected. However, the increase in overall performance illustrated in the last subsection could also lead to a boost in performance. Comparing the average number of correct sudokus solved by all possible tournament winners shows that in our experiment both effects are important. The average performance in the AA treatments was 2.93% lower than in the K treatment, although not significantly so. However, when controlling for age there is a statistically significant decrease of 6.46% for 4th graders and of 8.17% for 6th graders. Therefore, the decrease in the average performance of those selected as tournaments winners when AA is implemented is, although positive, not substantial.

3.5 The Effects of Affirmative Action on Expected Winning Probabilities

The implementation of AA might not only have a direct effect on performance but also might affect subjects' expectation about winning their respective tournament. This issue is of importance since it allows us to study whether affirmative action undermined self-confidence of the respective subjects. In question 6 of the questionnaire, subjects ranked their expectation of winning the tournament against their respective rival on an ordinal scale from 1 ("Definitely Not") to 5 ("Definitely"). As there was no information about the identity and characteristics of the respective opponent we use these answers as a measure of confidence in winning. PROB (1) and PROB (2) in Table 7 show ordered probit regressions using this measure of confidence in winning as dependent variable and "Pretest" and a treatment dummy for affirmative action ("AA" = 1 in treatments with affirmative action, ("AA" = 0 in treatments without it) as regressors. Quite intuitively, we find that high ability subjects in both the E and NE groups have higher confidence in winning their respective tournament as "Pretest" has a positive and significant coefficient at the 1% level in both regressions. More importantly, the implementation of AA does not significantly affect reported confidence of experienced subjects, while it significantly increases the confidence of non-experienced subjects at the 5% level. These results suggest that experienced subjects do not feel frustrated by the introduction of AA while, at the same time, its implementation correctly increases the expectations of non-experienced subjects to win their respective tournament.

Table 7: Expected Winning Probability, Affirmative Action and Ability		
	Experienced	Non-Experienced
	PROB (1) Dep. Var.: Win Prob.	PROB (2) Dep. Var.: Win Prob.
AA	-0.13 (0.16)	0.42 (0.21)**
Pretest	0.22 (0.06)***	0.12 (0.04)***
# Observations	179	148
Pseudo R^2	0.038	0.033

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** at the 1% level. Robust standard errors, clustered by treatment and class are in parentheses.

3.6 The Effects of Affirmative Action on Perception of Fairness

The fairness perception of different affirmative action policies might have important policy implications as fairness perceptions might determine the general acceptance of these policies and therefore the feasibility of their implementation. In fact, our analysis reveals that subjects perceive the different AA policies as substantially different with respect to their inherent fairness. We use the response to question No. 8 of the post-experimental questionnaire as a measure for fairness perception of the respective AA treatment. In this question subjects were asked for their perceived fairness of the implemented bonus in their treatment, where responses could vary between 1 (very fair) and 6 (very unfair).

For experienced subjects the responses are presented in Figure 4 and 5, where the AA treatments are pooled along types (LL&LL= “Lump-sum”, PL&PH= “Prop”) and sizes (LL&PL= “Low”, LH&PH= “High”). The figures suggest that experienced subjects perceived the high treatments as more unfair than the low treatments, and the lump-sum treatments as more unfair than the proportional treatments.

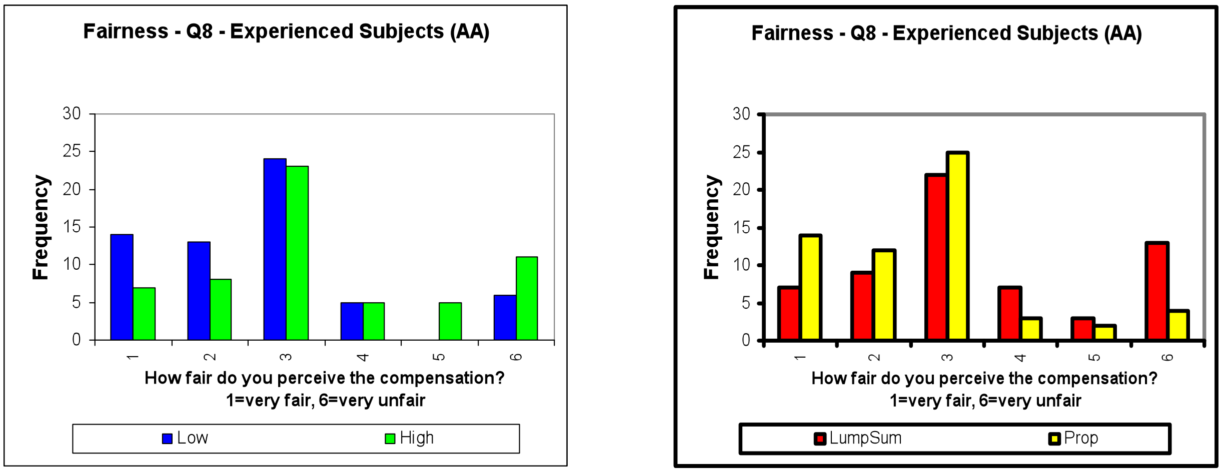


Figure 4: Fairness perception by experienced subjects for AA treatments.

The differences in fairness perceptions are confirmed by two-sided Mann-Whitney tests in the sense that the two distributions in each graph are significantly different from each other (p-value of 0.01 for low vs. high treatments and p-value of 0.01 for lump-sum vs. proportional treatments).[36] This is noteworthy given that on average the compensation received by subjects in the proportional treatments was higher (12 in the PL and 22 in the PH treatment) than those in the lump-sum treatments (8 in LL and 20 in LH). Hence, proportional types of AA (whose size depends on performance) seemed to be perceived as fairer by experienced subjects, although these subjects were on average handicapped more than by lump-sum types of affirmative action.[37]

For non-experienced subjects the implementation of affirmative action is generally perceived as being fairer than for experienced subjects. However, the relative fairness perception of different AA policies is quite similar among both subject pools. Figure 5 suggests that also for non-experienced subjects low (proportional) treatments seem to be perceived as slightly fairer than the high (lump-sum) treatments although non-experienced students actually benefit from high treatments. However, these distributions are not significantly different from each other,

[36] The LH treatment was perceived as being significantly more unfair than any other AA treatment (p-values of 0.04 for LL vs. LH, 0.01 for PL vs. LH, and 0.02 for PH vs. LH for a two-sided MW-test).
[37] Similar conclusions have been reported in the literature on positive fairness, see Konow (2000) for a survey.

which may be a result of non-experienced subjects being less able to assess the appropriateness of the compensation due to lack of exposure to the task.[38]

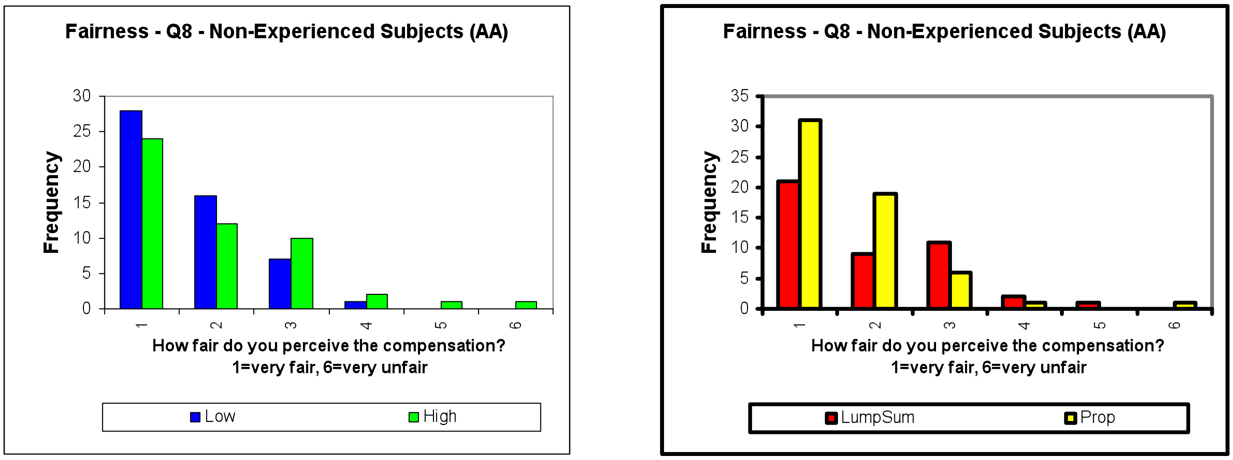
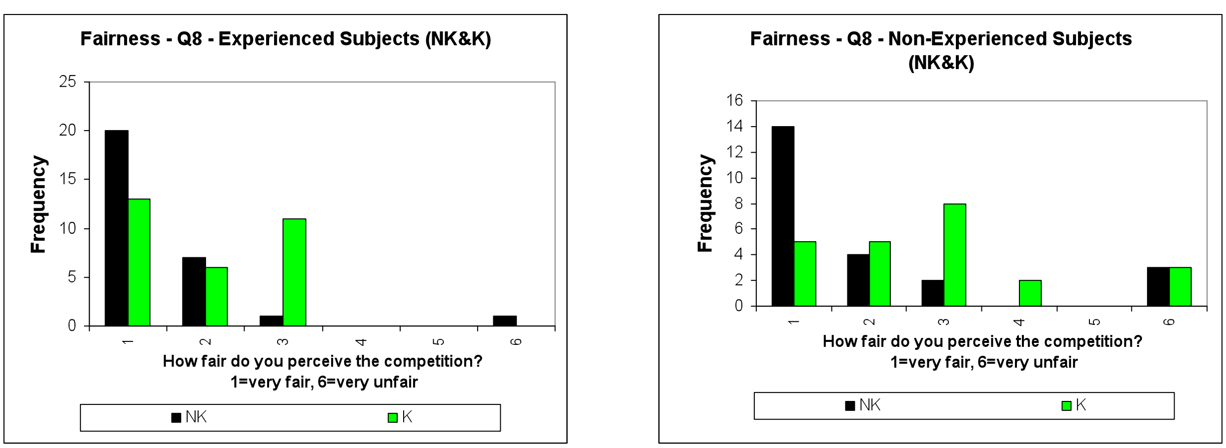


Figure 5: Fairness perception by non-experienced subjects for AA treatments.

Fairness perception of the benchmark treatments NK and K, where no affirmative action was implemented, can be used to evaluate the subjects' perception of competition under different information settings with respect to the differences in experience among opponents. For instance, experienced subjects perceived the tournament to be fairer in treatment NK when they were not aware of the asymmetry in experience in comparison to treatment K. This is suggested by the graph on the left side of Figure 6 and confirmed by a two-sided Mann-Whitney test (p-value of 0.02). Remarkably, subjects seem to evaluate the fairness of a treatment by abstracting, at least partially, from their self-interest.



[38] The only exception is again the LH treatment which is perceived as being significantly less fair than all other AA treatments (p-values of 0.07 for LL vs. LH, 0.07 for PL vs. LH, and of 0.06 for PH vs. LH for a two-sided MW-test).

Figure 6: Fairness perception by experienced (left) and non-experienced subjects (right) for K and NK

Results for non-experienced subjects point in the same direction. For instance, the fairness of treatment K is perceived as being significantly more unfair than treatment NK (p-value of 0.01 for two-sided MW-test), which is also suggested from the right part of Figure 6.

4. Conclusion

In this experimental study we take advantage of a situation where an exogenous disadvantage among two otherwise similar subject pools exists. This allows us to analyze the incentive effects of affirmative action in a real effort tournament setting where subjects have differences in capacities to compete with respect to the relevant task. Hence, the implementation of affirmative action can be evaluated without potential confounding effects due to self-selection, complex behavioural aspects like stereotype threats, or the potential artificiality of a laboratory. Additionally, our experimental design generates data on performance not only by selected tournament winners but by all participants.

Our empirical analysis of the experimental results suggests that the implementation of affirmative action leads to enhanced performance by a large fraction of participants and to a small decline in the average performance of selected winners, while balancing the pool of selecting winners. Our results also imply that different AA designs may significantly affect incentives and fairness perceptions. At the same time a large number of important questions have not been addressed in this study, for instance, the effects of affirmative action in situations where the disadvantage among subjects is less clear and more complex (like the mentioned stereotype threats or gender differences in propensities to compete), or the long run effects of these policies. We plan to address these issues in future research.

5. References

Balart, P., 2009. "Equality of Opportunities and Welfare in a Model of Job Allocation". Universitat Autònoma de Barcelona. Mimeo.

Bertrand, M., Hanna, R. and Mullainathan, S. “Affirmative Action: Evidence from College Admissions in India”, NBER Working Paper #13926. Forthcoming in *Journal of Public Economics*.

Calsamiglia, C., 2009. “Decentralizing Equality of Opportunity”. *International Economic Review* 50(1), 273-290.

Coate, S., and Loury, S., 1993. “Will Affirmative-Action Policies Eliminate Negative Stereotypes?”. *American Economic Review* 83(5), 1355-1360.

Chan, J., and Eyster, E., 2003. “Does Banning Affirmative Action Lower College Student Quality?”. *American Economic Review* 93(3), 858-872.

Che, Y., and Gale, I., 1998. “Caps on Political Lobbying”. *American Economic Review* 88(3), 643-651.

Franke, J., 2008. “Does Affirmative Action Reduce Effort Incentives? A Contest Game Analysis”. UAB and IAE Working Paper Series 711.07.

Freeman, R. B., Gelber, A. M., 2010. “Prize Structure and Information in Tournaments: Experimental Evidence”. *American Economic Journal: Applied Economics* (2), 149-164.

Fryer, R. G. J., and G. C. Loury, 2005. “Affirmative Action and Its Mythology”. *Journal of Economic Perspectives* 19(3), 147-162.

Fryer, R., and Loury, G. “Valuing Identity: The Simple Economics of Affirmative Action Policies”. Forthcoming in *Journal of Political Economy*.

Fu, Q., 2006. “A Theory of Affirmative Action in College Admissions”. *Economic Inquiry* 44, 420-428.

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

Gneezy, U., and A. Rustichini, 2004. "Gender and Competition at a Young Age," *American Economic Review, Papers and Proceedings*, 377-381.

Harbaugh, W., and Krause, K., 2000. "Children's Altruism in Public Good and Dictator Experiments," *Economic Inquiry* 38(1), 95-109.

Harbaugh, W., Krause, K. and Berry, T. R., 2001. "GARP for Kids. On the Development of Rational Choice Behavior". *American Economic Review* 91(5), 1539-1545.

Hoff, K., and P. Pandey, 2006. "Discrimination, Social Identity, and Durable Inequalities". *American Economic Review, Papers and Proceedings* 96(2), 206-211.

Holzer, H., and D. Neumark, 2000. "Assessing Affirmative Action". *Journal of Economic Literature* 38(3), 483-568.

Hickman, B., (2010). "Effort, Race Gaps and Affirmative Action: A Game-Theoretic Analysis of College Admissions". Mimeo.

Krasnokutskaya, E., and K. Seim, (2010). "Bid Preference Programs and Participation in Procurement". Forthcoming in *American Economic Review.*

Konow, J., 2003. "Which is the fairest one of all? A positive analysis of justice theories". *Journal of Economic Literature* 41(4), 1186-1237.

Lazear, E..P., and S. Rosen, 1981. "Rank-Order Tournaments as Optimum Labor Contracts". *Journal of Political Economy* 1989(5), 841-864.

McCrary, J., (2007) "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police". *American Economic Review* 97(1), 318-353.

Marion, J., (2009). “How Costly Is Affirmative Action? Government Contracting and California's Proposition 209”. *The Review of Economics and Statistics* 91(3), 503-522.

Miller, A., and Segal, C., 2008. “Do Temporary Affirmative Action Programs Produce Permanent Effects? A Study of Minority and Female Employment in Law Enforcement Agencies”. Universitat Pompeu Fabra. Mimeo.

Myerson, R., 1981: “Optimal auction design”. *Mathematics of Operations Research* 6, 58-73.

Niederle, M., C. Segal and Vesterlund, L., 2008. “How Costly is Diversity? Affirmative Action in Competitive Environments”. Universitat Pompeu Fabra. Mimeo.

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

Schotter, A., and Weigelt, K., 1992. “Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results”. *Quarterly Journal of Economics* 107(2), 511-39.

6. Appendix: Experimental Instructions

Below you can find a translation of the experimental instructions used in the experiment. Instructions for all treatments and schools were identical apart from the changes here indicated. Sentences in bold were not included in the NK treatment. The sentences in *bold and italics* were only included in the treatments with affirmative action (LH, LL, PH, PL). Words in (parentheses) indicate changes between the experiences and non-experienced schools and changes in the type of the compensations (lump-sum or proportional). Sizes of the compensations varied as explained in section 2. Numerical examples varied in order to reflect changes in compensation sizes, but were created such as the results of both contestants were the same. A whole set of instructions is available upon request. Instructions were originally written in Spanish.

Pre-instructions

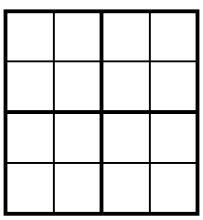
Your Code: ____

Thank you for participating. First, we are going to explain what you will be doing.

You have to fill in grids with the numbers 1, 2, 3 and 4.
To do this you have to use the following rules:

1. All boxes in a grid must be filled in with a number.
2. The same number can appear only once in each column (vertical).
3. The same number can appear only once in each row (horizontal).
4. The same number can appear only once in each square. Each grid is divided in 4 squares, marked in bold lines.
5. In each grid all numbers 1, 2, 3 and 4 must be in each column, each row, and each square.

Here are some examples:



This column is **completed wrongly** because the 3 appears twice (rule 2)

This column is **completed correctly.**

This row is **completed wrongly** because the 4 appears twice (rule 3

This row is **completed correctly.**

This square is **completed wrongly**

This square is **completed correctly.**

because the 1 appears twice (rule 4)

This is an example of a correctly completed grid.

Before starting you have 5 minutes to complete the following grids to check whether you have understood the rules. We will give you the correctly completed grids after the 5 minutes period.

Please remain silent and on your seat without disturbing anyone during the whole practice.

Raise your hand after you have finished all grids and we will pick them up.

Good luck!

Instructions

Your Code: ____

You are randomly matched with another student, your matched participant, from another school similar to yours, who is completing the same grids as you are.

The students at the other school have (NOT) learned before how to solve these types of grids because it was (NOT) taught to them in their math classes.

You have now 30 minutes time to complete as many grids as possible with the numbers **1, 2, 3 and 4** on the formulaires that we are now going to distribute.

We will compare how many grids you have solved correctly with the number of correctly solved grids by your matched participant from the other school:

- If you have correctly solved more grids then you will earn a 7 EU voucher that you can redeem in “La Casa del Libro”, where you can buy books, collector’s cards, toys, music or comics.
- If you have correctly solved less grids then you will not earn the voucher.
- If you have correctly solved the same number of grids, then a toss of a coin will be used to determine who earns the voucher.

To compensate (the other students) for the fact that (they)/(you) have (less)/(more) practice (than you) we are going to give (them)/(you) (20 extra grids)/(1 grid more for each grid that (they)/(you) solve correctly).

For example *(example provided for the PH Treatment)*:

- If your matched participant correctly solves **12** grids, they count as 12 +12 = 24 grids. Therefore you will earn the voucher if you solve correctly 25 grids or more.
- If your matched participant correctly solves **30** grids, they count as 30 +30 = 60 grids. Therefore you will not earn the voucher if you solve correctly 59 grids or less.
- If your matched participant correctly solves **20** grids, they count as 20 +20 = 40 grids. Therefore, if you solve correctly 40 grids, a toss of a coin determines whether you earn the voucher.

The numbers of this example are chosen randomly and do not indicate how many grids a student can solve correctly.
We would like to inform you that we have studied the results of other students of your age from other schools who completed the same grids: The maximum number of grids that somebody managed to solve correctly in 30 minutes were 81 grids and the minimum was 0 grids. On average the students completed around 25 grids correctly.
Remember that only correctly solved grids count.
Wait to turn the answer sheet until we tell you to do so. You have 30 minutes. Good luck!

Your Code: ____

Thank you for your participation.

Final Questionnaire

Please answer the following questions:

1. How did you find today's task?

   Interesting Entertaining A bit long Boring

2. How many grids like these have you tried before?

   None Between 1 and 5 Between 6 and 20 Between 20 and 40 More than 40

3. If you have tried solving grids like these before, where did you do it? _________

4. How many grids do you think you have solved correctly today? ______________

5. How many grids do you think your partner of the other school has solved correctly?____

6. Do you think you are going to get the voucher?

   Definitely Probably yes I don't know Probably not Definitely not

7. Do you think it was a good idea to compensate the students of the other school that did not do grids like this before in school?

   YES NO

8. The competition with the students of the other school from my perspective seemed to be:

   Fair Rather Fair A bit Unfair Unfair Rather Unfair Very Unfair

9. Any other comment? ____