The Lifecycle of Affirmative Action Policies and Its Effect on Effort and Sabotage Behavior

Subhasish M. Chowdhury (University of Sheffield)
Anastasia Danilov (HU Berlin)
Martin G. Kocher (University of Vienna, University of Gothenburg, CESifo Munich)

Discussion Paper No. 401

June 14, 2023

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

The Lifecycle of Affirmative Action Policies and Its Effect on Effort and Sabotage Behavior*

Subhasish M. Chowdhury [a], Anastasia Danilov [b], and Martin G. Kocher [c]

[a] Department of Economics, University of Sheffield, Sheffield S1 4DT, UK.

[b] School of Management and Economics, Humboldt-University of Berlin, 10178 Berlin, Germany.

[c] Department of Economics, University of Vienna, 1090 Vienna, Austria; Department of Economics, University of Gothenburg, Gothenburg, Sweden; and CESifo Munich.

This version: May 16, 2023

Abstract

A main goal of affirmative action (AA) policies is to enable disadvantaged groups to compete with their privileged counterparts. Existing theoretical and empirical research documents that incorporating AA can result in both more egalitarian outcomes and higher exerted efforts. However, the direct behavioral effects of the introduction and removal of such policies are still under-researched. It is also unclear how specific AA policy instruments, for instance, head-start for a disadvantaged group or handicap for the privileged group, affect behavior. We examine these questions in a laboratory experiment in which individuals participate in a real-effort tournament and can sabotage each other. We find that AA does not necessarily result in higher effort. High performers that already experienced an existing AA-free tournament reduce their effort levels after the introduction of the AA policy. Additionally, we observe less sabotage under AA when the tournament started directly with the AA regime. The removal of AA policies, however, significantly intensifies sabotage. Finally, there are no overall systematic differences between handicap and head-start in terms of effort provision or sabotaging behavior.

JEL Classification: C72; C91; D63; D72

Keywords: Affirmative action; Sabotage; Experiment; Tournament; Handicap; Head-start

---

* Corresponding author: Anastasia Danilov (anastasia.danilov@hu-berlin.de).

We thank Florian Englmaier, Ben Greiner, Wieland Müller, Marie Claire Villeval, the seminar and workshop participants at Seoul National University, the IIM Calcutta, the Vienna Center for Experimental Economics, the Micro and Macro Foundation of Conflicts Conference at Bath, the ESA World Congress at Berlin, the Workshop in Honor of Vernon Smith at Dijon, the SFB Annual Meeting in Schwanenwerder, the Reading Experimental and Behavioral Economics Workshop, and the IAAEU "Workshop on Organizations, Incentives and Performance Pay for their very helpful comments. Ethical approval for the study was obtained at the University of Bath (EIRA1-5352). We gratefully acknowledge the financial support by the German Research Foundation through CRC 'Rationality and Competition' (TRR 190 Project number 280092119) and research unit 'Design and Behavior' (FOR 1371). Any remaining errors are our own.

1. Introduction

Tournaments and contests are used widely in various organizational, political, and social domains. People participate in several contests such as hiring decisions, promotions, college admissions, grant applications, sport tournaments, and procurement auctions (Balafoutas et al., 2019; Konrad, 2009). In these situations, the winners of a contest are selected based on their relative performance, which is determined by a combination of costly efforts exerted by the contestants, their ability levels and, potentially, a luck or random component. The standard theoretical predictions suggest that relative rewards such as in tournaments often offer high-powered incentives and lead to high overall effort levels being provided. In line with these predictions, the designers of such (promotional, sports, or funding) tournaments are often interested in achieving higher overall effort provision.

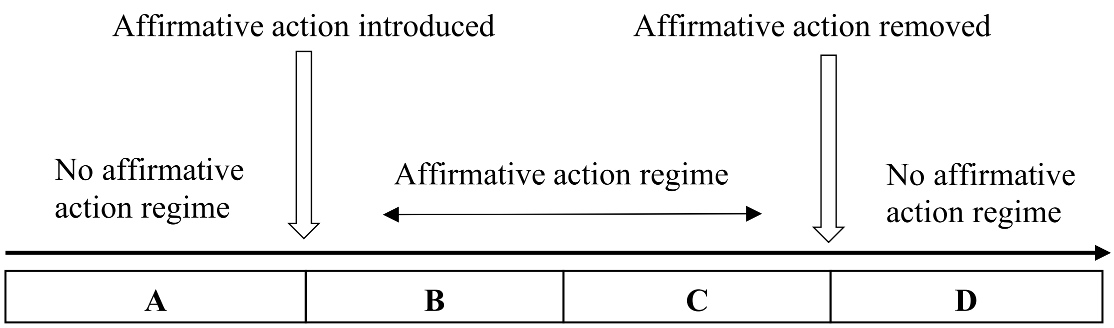
In many tournaments, however, the contestants exhibit very different ability levels. This implies that one (or a group) of the contestants has a-priori higher chances of receiving the tournament prize than others (in the following also termed 'favorites' and 'underdogs'). Examples include higher ranked players in sport competitions, researchers with better records of publications in grant applications, privileged students in college applications, more qualified job seekers in the job market, etc. (Fu, 2006; Franke, 2012). Such heterogeneous competitions fail to provide an even-level playing field for the left-behind groups and may result in undesirable outcomes, such as low effort, reduced contest participation, exacerbating income inequality, and lack of diversity, to name a few (Chowdhury et al., 2023). A designer that is either concerned about maximizing effort or aiming for diversity in the tournaments (e.g., in sports or in the workplace) often employs various instruments to level the playing field, known as 'competitive balance' (Fort and Maxcy, 2003) in sports. When introduced due to observable characteristics that cannot be changed by the individual such as gender, racial, ethnical, or socio-economic background, such policies are referred to as *affirmative action* (Holzer and Neumark, 2000).

There are a variety of affirmative action (AA) instruments in practice. The three most common instruments are (i) handicap, in which favorites are weakened a priori; (ii) head-start, in which underdogs are strengthened a priori; and (iii) quota, in which some winning prizes are reserved for the underdogs.

Politically, the idea of AA has always been controversial. Advocates of AA claim that it, among others, reverses historical injustice, counterbalances the difference in abilities, helps to achieve a more egalitarian outcome, as well as induces higher effort levels (Schotter and Weigelt, 1992). Reflecting this, US President Lyndon B. Johnson famously argued in 1965 that "*You do not take a person who for years has been hobbled by chains and liberate him, bring him up to the starting point of a race and then say you're free to compete with all the others*". Opposition to AA raises the issue that it is a type of (reverse) discrimination for those that are not supported, it might lead to inefficiencies, and it can reduce total effort provided.

Furthermore, in many tournaments in which relative performance determines the final outcome players cannot only exert effort to improve their own performance, but they can also exert effort to reduce the performance of the rivals (Konrad, 2000; Chowdhury and Gürtler, 2015). Such acts of *sabotage* include spreading rumors, withholding information, damaging outputs, etc. Sabotage is undesirable for most contest designers, but it is especially harmful in the settings of workplace and organizations. Any competitive balance policy that can affect effort provision can potentially also affect sabotage behavior. Hence, when it is possible for contestants to sabotage their competitors, the gross effect of AA as a combination of productive effort and destructive effort becomes relevant.

FIGURE 1. THE LIFECYCLE OF AN AFFIRMATIVE ACTION POLICY



Notably, a possible 'lifecycle' of AA, with AA policies being introduced or removed along the way, may induce additional behavioral responses that have not been studied well in controlled experiments so far (see Chowdhury et al., 2023). A possible lifecycle starts with a tournament without an AA policy in place. Then, the designer might introduce an AA policy, and the contestants could change their behavior to adjust to it. After some time, the designer might want to remove the AA policy, and contestants again might adjust to the new regime. Such dynamic effects depicted conceptually in Figure 1 have not been investigated in a rigorous economic experiment, yet.

The longer horizontal arrow illustrates a potential timeline, and the lifecycle of AA can be divided analytically into four phases, following an overlapping generation logic. Initially, there is no AA policy in phase A. Then, the AA policy is introduced, and the time the AA regime is in place can be broken down analytically into two phases: B and C. In phase B, contestants were 'born' in phase A and will now have to adjust to the new regime with AA. However, after some time, other contestants are 'born into' the AA regime in phase C who did not previously experience a regime without AA. Phase D depicts the situation after the removal of the AA policy. Note that contestants in phase D that were 'born' in an earlier phase with AA will potentially have to adjust to the change in policy.

To investigate the immediate effects of the introduction of AA policies in terms of effort provision and sabotaging one must compare behavior in phase A and phase B. Comparing behavior in phase C and phase D will show the effects of the removal of such a policy. Moreover, both phases A and D are regimes without AA. However, contestants in phase A do not have any prior experience of an AA policy, whereas their counterparts in phase D do. Similarly, both phases B and C are regimes with AA. However, contestants in phase C do not have any prior experience of a tournament regime without AA, whereas people in phase B do.

It is extremely hard to rigorously identify behavioral responses in such dynamics in the field, outside the experimental laboratory (Schotter and Weigelt, 1992). Therefore, we investigate effort exertion and sabotaging behavior in real-effort laboratory tournaments in which contestants of heterogeneous abilities compete for monetary rewards repeatedly. Subjects have a tangible option to sabotage each other. We study the effects of the introduction and removal of two specific ability-based AA instruments: head-start (underdogs are strengthened a priori), and handicap (favorites are weakened a priori). We hypothesize based on a theoretical model laid out in Section 3 that the introduction of any type of AA policy should result in higher effort levels, but also in more sabotage. Furthermore, we hypothesize that the removal of the AA instrument will bring back the status quo ante in terms of effort provided and sabotaging – independent of the policy or the decision-maker type.

Our empirical results indicate that neither favorites nor underdogs increase their effort significantly after the introduction of any AA policy. Furthermore, the increase in sabotage under the AA policy turns out to be less pronounced than expected; for head-start, sabotaging behavior even is reduced. On the other hand, the removal of AA policies results in higher efforts of favorites and more sabotage from both types. We then identify the conditions under which AA policies work as an effective instrument to propel egalitarian outcomes and increase effort exerted by underdogs. This occurs when contestants are 'born into' an AA regime, i.e., they do not have any previous experience of a competitive environment without AA. Notably, we do not observe any significant differences in the effects of head-start and handicap on effort provision and sabotage. Therefore, we conclude that (i) the contestants' previous experiences with the competitive environment plays an important role in shaping their reactions towards AA policies; and (ii) AA does not induce more sabotaging when it is introduced, but interestingly, we find some evidence that sabotaging is more prevalent after AA policies have been removed.

This study contributes to the literature on AA in two relevant aspects. It is the first experiment to test and compare the effects of different types of AA instruments within a coherent experimental framework. It is also the first experiment to systematically investigate the policy dynamics (introduction and removal) of AA policies – providing insights to scholars in contest research and policy makers alike. Moreover, the current study delivers new insights in the context of the literature on sabotage behavior in a real effort setting by interacting such behavior with (AA) policy instruments.

The rest of the paper is structured as follows. Section 2 discusses the existing literature. In Section 3, we provide the theoretical background and hypotheses. The experimental design is described in Section 4, and the structure of the data and the identification strategy are presented in Section 5. Sections 6 and 7 report our results on effort and on sabotage. Finally, Section 8 concludes.

2. Literature Review

As discussed, the research scope of the current study, broadly speaking, covers three related areas of the literature: contests, AA, and sabotage. Contests are games in which players expend costly resources (effort) to win a prize. Some prominent examples – as provided in the introduction – are promotional tournaments, grant applications, sports tournaments, etc. An excellent overview of the contest literature can be found in the book by Konrad (2009) or in a recent survey by Corchón and Serena (2018). Many of the contests (e.g., job interviews, promotions, or sport competitions) are created by a contest designer with specific objectives. It is noted in the literature (Konrad, 2009; Chowdhury et al., 2023) that in organizational contests – the contests that we are focusing on in our study – the designer is often interested in maximizing the total effort exerted. Various studies (e.g., Che and Gale, 2003; Moldovanu and Sela, 2001, 2006; Fu and Lu, 2009; Chowdhury and Kim, 2017, among many others) analyze how various contest rules, cost structure of effort, level of randomness in the effort-outcome relationship, etc. contribute to such a goal. For the interested reader, Chowdhury et al. (2023) provide a comprehensive survey of heterogeneity and AA in contests. We contribute to this area of literature experimentally by showing how various AA policy instruments and a sabotage option can affect overall effort and performance.

Out of the various policy measures in contests, we focus on the instruments of AA that are usually introduced to level the playing field when the contestants exhibit heterogeneous ability levels ex ante, either due to innate differences or due to historical injustice. The impact of these instruments has been broadly analyzed in several disciplines. In the following, we concentrate on the AA literature in the context of contest theory. Early theoretical studies point out the positive effects of AA. Fryer and Loury (2005) find that profile-specific AA can increase effort and reduce inequality. Fu (2006) shows that such policies may improve academic test scores when admitting new students. Similar results are obtained when employing various contest structures, number of players, and information settings in the models (Franke, 2012; Franke et al., 2013; Lee, 2013; Calsamiglia et al., 2013). Another set of studies (Fain, 2009; Kirkegaard, 2012; Krishna and Tarasov, 2016; Dahm and Esteve-González, 2018) lay down a variety of mechanisms for which an AA policy can enhance effort by considering issues such as inequality and contest participation. Testing such theories, Schotter and Weigelt (1992) show experimentally that these policies benefit the disadvantaged group and increase the effort levels of all contestants. Balafoutas and Sutter (2012) provide experimental evidence that employing AA to level the playing field for female contestants improves female participation, but exerted effort levels remain the same. Czibor and Martinez (2019) find a positive effect of AA on women's willingness to compete. We contribute to this literature by introducing a new experimental paradigm that compares two AA policies in the laboratory and by investigating the introduction and removal of such policies in the laboratory for the first time.

Finally, we also contribute to the literature on sabotage as a deliberate act of damaging an opponent's effort or output to improve one's own relative performance. The idea was first introduced in the contest literature by Lazear (1989). The theoretical literature was developed later by many scholars (Konrad, 2000; Chen, 2003; Kräkel, 2005; Amegashie and Runkel, 2007; Münster, 2007; Gürtler, 2008; Soubeyran, 2009; Gürtler and Münster, 2010, 2013, among others). There is also a growing number of experimental research from the laboratory on sabotage behavior in the context of contests (Harbring et al., 2007; Harbring and Irlenbusch, 2005, 2008, 2011; Carpenter et al., 2010; Danilov et al., 2019; Dato and Nieken, 2019). In addition, several field studies (del Corral et al., 2010; Deutscher et al., 2013; Garicano and Palacios-Huerta, 2014) identify sabotage behavior in contests. Closer to our specific interest, Brown and Chowdhury (2017) show with horse-racing data that handicapping increases sabotage behavior among jockeys. Steinmayr et al. (2018) document more egalitarian outcomes in balanced swimming relays with appropriately chosen handicap and head-start. Leibbrandt et al. (2018) find experimentally that introducing gender quotas may increase distorted peer reviewing against women, mostly done by women. Fallucchi and Quercia (2018) provide evidence that introducing an AA policy increases retaliation against the designer. Petters and Schröder (2020) report that quotas intensify sabotage that targets the advantaged types by the disadvantaged types. The surveys by Chowdhury and Gürtler (2015) and Piest and Schreck (2020) give comprehensive reviews on the effects of sabotage in contests.

As already mentioned, the literature so far, however, is silent on the interaction of different types of AA policies with sabotage. Testing their potential effects empirically is important because there might be behavioral effects that are not accounted for in theoretical models using standard assumptions. For instance, whereas a head-start a priori may not trigger negative emotions due to its positive frame among those who are not supported, a handicap may do so. Moreover, the existing literature on the effects of the lifecycle of AA policies is very rare: whether an introduction or removal of such policies has effects on subsequent effort levels and sabotage behavior through behavioral spillovers of the previous experience, is a relevant aspect for a contest designer. Our study focuses on these two aspects.

3. Theoretical Benchmark and Hypotheses

We consider a two-player tournament with sabotage (Lazear and Rosen, 1982; Lazear, 1989) in which player $i \in \{1,2\}$ can exert costly effort $e_i \in \mathbb{R}_+$ to enhance own performance, or costly sabotage $s_j \in \mathbb{R}_+$ to impede the effort of the opponent j (where $i \neq j$). Following the standard structure in the literature (e.g., Gill and Stone, 2010; Brown and Chowdhury, 2017) we denote the 'output' of player i as:

$$y_i = \bar{a}_i + a_i + e_i - \alpha s_i + \varepsilon_i$$

where $\bar{a}_i \in \mathbb{R}_+$ is the ex-ante efficiency or ability level, $a_i \in \mathbb{R}$ is the AA policy introduced by the contest designer towards player i, $\alpha \in (0,1)$ is a parameter, s_i is the sabotage inflicted on player i by player j, and $\varepsilon_i \in \mathbb{R}$ is a noise term with known distribution.

The players simultaneously and independently exert efforts and commit sabotage, and the player with the highest output wins a prize of common value $v > 0$. In the case of a tie, the prize is given to either player with the same likelihood. Hence, the Contest Success Function is:

$$p_i = \begin{cases} 1 & if\ y_i > y_j \\ 1/2 & if\ y_i = y_j. \\ 0 & if\ y_i < y_j \end{cases}$$

Then, the probability that player i wins is:

$$p_i = \text{Prob}(y_i > y_j) = \text{Prob}\left((\bar{a}_i + a_i + e_i - \alpha s_i + \varepsilon_i) > (\bar{a}_j + a_j + e_j - \alpha s_j + \varepsilon_j)\right)$$

$$= \text{Prob}\left((\bar{a}_i + a_i + e_i - \alpha s_i) - (\bar{a}_j + a_j + e_j - \alpha s_j) > (\varepsilon_j - \varepsilon_i)\right)$$

$$= G\left(\Delta a_i + (e_i - e_j) - \alpha(s_i - s_j)\right)$$

where $\Delta a_i = (\bar{a}_i + a_i - \bar{a}_j - a_j)$, and $G(.)$ is the CDF of $(\varepsilon_j - \varepsilon_i)$ with unimodal PDF $g(.)$.

Players face the common cost function $c_i = c(e_i, s_j)$ with the standard properties: $c(0,0) = 0$, $\frac{\partial c}{\partial e_i} > 0, \frac{\partial c}{\partial s_j} > 0, \frac{\partial^2 c}{\partial e_i^2} \geq 0, \frac{\partial^2 c}{\partial s_j^2} \geq 0, \frac{\partial^2 c}{\partial e_i \partial s_j} \geq 0$.

Hence, the payoff function of player i, $\pi_i = p_i v - c_i$, can be rewritten as:

$$\pi_i(e_i, s_j) = G\left(\Delta a_i + (e_i - e_j) - \alpha(s_i - s_j)\right)v - c_i(e_i, s_j) \tag{1}$$

Solving for the first order conditions (FOCs) of (1) we get:

$$g\big(\Delta a_i + (e_i - e_j) - \alpha(s_i - s_j)\big)v = \frac{\partial c}{\partial e_i}$$

$$g\big(\Delta a_i + (e_i - e_j) - \alpha(s_i - s_j)\big)v = \frac{\partial c}{\partial s_j}$$

Comparing the FOCs of the two players, we observe that $\frac{\partial c}{\partial e_1} = \frac{\partial c}{\partial e_2}$ and $\frac{\partial c}{\partial s_2} = \frac{\partial c}{\partial s_1}$. This implies that there is a symmetric equilibrium for which $e_1^* = e_2^*$ and $s_1^* = s_2^*$.

Note that for $e_1^* = e_2^* = e^*$ and $s_1^* = s_2^* = s^*$, the FOCs reduce to $g(\Delta a_i)v = \frac{\partial c(e^*,s^*)}{\partial e_i}$ and $g(\Delta a_i)v = \frac{\partial c(e^*,s^*)}{\partial s_j}$. Without loss of generality, define $(\bar{a}_1 - \bar{a}_2) > 0$ as the ex-ante ability difference, i.e., Player 1 is the 'favorite' and has a higher ex-ante likelihood of winning the tournament, whereas Player 2 is the 'underdog'.

An AA policy is introduced in a way such that the designer either adds or subtracts to the effort of the players: i.e., $a_1 < 0$ or $a_2 > 0$. In effect, $\frac{\partial |\Delta a_i|}{\partial |a_1|} < 0$ and $\frac{\partial |\Delta a_i|}{\partial a_2} < 0$, i.e., both a_1 (handicap for the favorite) and a_2 (head-start to the underdog) reduces the a-priori asymmetry in ability. Given the unimodal shape of the PDF, an AA policy will effectively increase $g(\Delta a_i)v$. To balance the FOC, the marginal costs must increase; and from the convexity of the cost functions, this means both equilibrium effort (e^*) as well as equilibrium sabotage (s^*) should increase.

The simple model based on standard preference assumptions provides us with a set of theoretical predictions that we introduce below. We start with the overall effects of the AA policies as delineated above:

Hypothesis 1. The introduction of an AA policy (either head-start or handicap) will increase (i) effort and (ii) sabotage for both the favorite and the underdog.

The model can also be used to make predictions for the removal of AA policies: ceteris paribus, when an existing AA policy is removed, then both effort and sabotage return to their 'original' levels, i.e., effort and sabotage will both decrease. Obviously, in the standard model, there are no spillovers. Furthermore, this theoretical result is invariant in the player's identity. This provides us with our second hypothesis:

Hypothesis 2. The removal of an existing AA policy (either head-start or handicap) will decrease (i) effort and (ii) sabotage for both the favorite and the underdog.

Further, note that this theoretical effect is independent of the nature of the policy as well. That is, if $\frac{\partial|\Delta a_i|}{\partial|a_1|} = \frac{\partial|\Delta a_i|}{\partial a_2}$, then $\frac{\partial e^*}{\partial|a_1|} = \frac{\partial e^*}{\partial a_2}$ as well as $\frac{\partial s^*}{\partial|a_1|} = \frac{\partial s^*}{\partial a_2}$. This gives our next hypothesis:

Hypothesis 3. If the reduction in heterogeneity is the same for handicap and head-start, then the effect on effort and sabotage will be the same for the two policies.

We test the three hypotheses in a laboratory experiment. Before we do so, it is important to mention explicitly that there can be behavioral effects not captured by standard preference or standard rationality assumptions. First, inertia in action due to habit formation may occur. As a result, even after an AA policy is removed, effort and sabotage levels may stay at higher levels than predicted by standard theory. This type of behavioral inertia is observed in other contexts such as pricing behavior (e.g., Chowdhury and Crede, 2020). Obviously, a policymaker that is required to remove the policy would hope for habit formation only in effort, and not in sabotage. Second, there can be non-invariance of player reactions in the sense that favorites may react differently than underdogs when a policy is implemented or removed. Whereas the AA policies help the underdogs in the contest, it worsens the relative position of the favorite in the contest. Hence, it may result in either a discouragement effect in terms of effort or trigger spiteful behavior in terms of sabotage rather among the favorites than the underdogs (Fallucchi and Quercia, 2018; Girard, 2018). Finally, whereas head-start provides support to the underdog, handicap impedes the possibilities of the favorite. Hence, there may be non-invariance of AA policy choices due to contestants' different perceptions of handicap and head-start policies, following specific interpretations of procedural fairness concerns (Martin et al., 2020). For example, contestants might increase effort less and increase sabotage more under a handicap policy than under a head-start policy, because they perceive handicap as less fair than head-start. Over the periods, the underdogs will adjust their effort provision, reacting to the favorite's behavior. Hence, overall head-start could produce more effort and less sabotage than handicap.

4. Experimental Design and Procedures

We conducted a laboratory experiment at the University of Cologne in which a total of 192 subjects took part. The subjects were students of various study disciplines, recruited through the ORSEE lab management software (Greiner, 2015). The average age of the subjects was 24.4 years and about half (47.4%) were females. As detailed below, we employed four between-subjects treatments with two sessions for each. In each session, there were 24 subjects. The experiment was computerized and coded with the help of the z-tree software (Fischbacher, 2007). At the beginning of the experiment, all subjects received general written instructions informing them that the experiment consists of four parts and that there are several identical rounds in Part 1, 2, and 3. Subjects learned that their final earnings would be the sum of the results from one randomly chosen round from each of these three parts, plus earnings from Part 4, and a €4.00 show-up fee. On average, sessions took about an hour and forty-five minutes, and the average earning was €18.40.

The specific instructions regarding the content of an experimental part were provided at the beginning of each part.[1] Part 1 included an individual working phase with eight rounds of two minutes each. The underlying task was to answer as many math questions as possible, with a monetary incentive for each correct answer. According to literature, we call this the 'piece rate' mechanism. In both Parts 2 and 3, subjects were matched into pairs and competed for a winner-takes-all tournament prize for working on a different set of math questions (of a similar nature as to those in Part 1). Each of these two parts consisted of eight two-minute rounds, described in more detail below. Part 4 involved a one-shot gamble measuring subjects' risk attitudes, as developed by Eckel and Grossman (2008). Concluding Part 4, we ran a survey to collect data on demographic characteristics of subjects. At the end of the experiment, one round was randomly chosen for payment from each of the first three parts. Together with the earnings from Part 4, subjects were privately paid in cash at the very end of the experiment. Further details about the real-effort task, individual working phases, tournament procedures and the possibility to sabotage are delineated below.

[1] After subjects had read instructions, they had a chance to (privately) ask questions. After all questions had been clarified, subjects were required to complete a comprehension quiz. Only after that could they proceed with the experiment. Complete instructions and questionnaires can be found in Online Appendix III.

4.1. The Real-Effort Task

The real-effort task used in our experiment was inspired by Dohmen and Falk (2011). Subjects had to perform simple arithmetic computations such as additions, subtractions, multiplications, or divisions of two one- or two-digit numbers and enter their answers on the computer screen. A pencil and some papers were provided, but no calculators or other tools (smartphones) were allowed. Each round lasted for two minutes and had a total pool of 50 unique questions (different in each round). Subjects did not know the total number of available questions. The maximum number of correctly solved questions was 43, suggesting that the question pool of 50 was large enough even for the most productive subject.

All subjects worked on the same math questions. Even though the questions were different in every round, we aimed to keep the difficulty of the questions similar between rounds. The question pool in each round included five very easy computations (level one, e.g., 6*6=?, 14-5=?), and 15 computations each from level two (e.g., 57-12=?, 5*21=?), level three (e.g., 3*41=?, 72/6=?) as well as level four (e.g., 7*61=?, 11*24=?). The order of questions was randomized in a way such that the sequence of difficulty levels was the same in each round. Subjects had no information about the question pool composition.

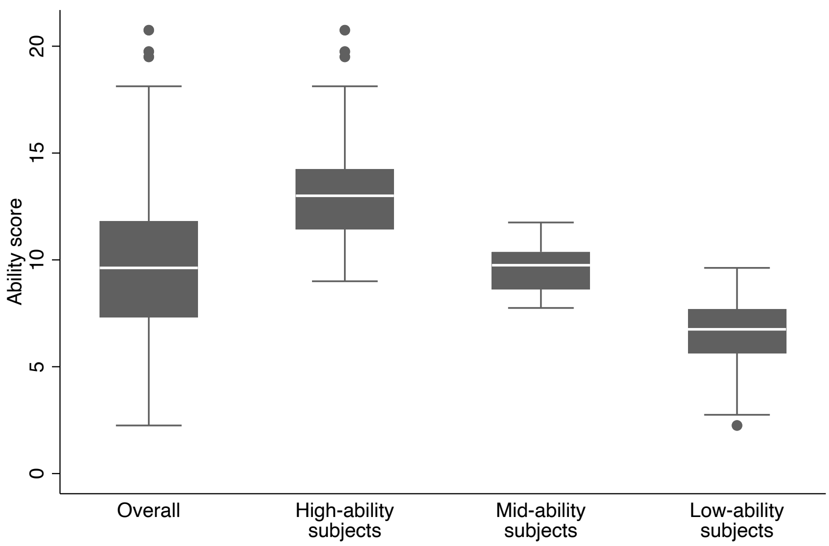
Instant feedback about the correctness of their answers was given to subjects, and they could always see their scores on the screen. As soon as subjects entered an answer and clicked 'OK', a new question appeared. In case the answer was correct, the score increased by one point. Correctly solved questions were not asked again. If the answer was wrong, the score remained unchanged, and the question could be asked again later in the same round. In this manner, we attempted to provide subjects an opportunity to work at their own pace and continue computations even if they were not able to solve a particular calculation.[2]

[2] We are aware of the possibility that subjects could avoid more difficult questions by submitting random answers. Nevertheless, we believe that the chance of an intentional selection of question is not essential for our results, because: (i) the number of easy questions was quite low (five), and it was necessary to answer more difficult questions to achieve a sufficiently high score; (ii) the unanswered questions were placed 'back into the loop' and asked again; (iii) subjects did not know how difficult subsequent questions would be; and (iv) since all subjects in all treatments went through the same questions in the same order, everyone had an equal opportunity to 'pick' an easy question if they intended to do so. For these reasons, we believe to have sufficient control over the possible intentional skipping of difficult questions.

4.2. Measure of Individual Ability

Part 1 was identical in all treatments. It consisted of eight rounds of two minutes each, where subjects worked individually on the real-effort task described above. At the end of the experiment, one of the rounds was randomly selected, and subjects received 5 ECU (€0.15) for each correctly answered question in this round.

FIGURE 2. ABILITY SCORES BY CATEGORY



Notes: Box-and-whiskers plots of ability scores measured as the average number of correct answers per minute in rounds five to eight of Part 1.

Based on the data in Part 1, we computed mean individual scores (i.e., the number of correctly solved questions) and used them as individual proxies of subjects' abilities. Due to a steep learning curve at the beginning of the experiment, the average number of correct answers goes up by 27.3% in the second half of Part 1 as compared to its first half ($p < 0.01$, two-sided Wilcoxon signed-rank test; see also Table A1 in Appendix I). Thus, our computation of ability score is based only on the second half of Part 1, i.e., rounds five to eight. The relative ranking of ability scores was used to classify subjects into either Category I (favorite or 'high-ability', 37.5% of 24 subjects per session), Category II ('mid-ability', 25%), or Category III (underdog or 'low-ability', 37.5%). Figure 2 illustrates the distribution of ability scores for these categories. The mean ability score of high-ability types amounts to 13.28 correct answers per minute and is 39.2% higher than the ability scores of mid-ability types (9.54), and more than twice as high as of low-ability types (6.62). The differences are highly significant ($p < 0.001$

for pairwise comparisons; two-sided Fisher-Pitman permutation tests for two independent samples, using Monte-Carlo simulation with 200,000 runs – henceforth FP2S test).[3]

4.3. Tournament with or without Affirmative Action Policies

In Parts 2 and 3, subjects compete in a tournament according to the following protocol: at the beginning of Part 2, subjects learned that they were divided into three ability categories (I, II, and III), based on their relative performances in Part 1, according to the procedure described above. However, they did not know the number of subjects in each category or any performance thresholds for the classification. Then, subjects were matched into pairs and received information about their own category and the category of their partner. All subjects of the high ability type (Category I) were randomly paired with one of the subjects of the low ability type (Category III). The subjects of the mid-ability type were randomly paired with each other. This categorization and the pair composition remained unchanged for all rounds of both parts. In this paper, we will focus solely on heterogeneous pairs that consisted of one high-ability and one low-ability type contestant. Before the first round of Part 3, subjects were informed that they will interact with the same partner as in Part 2. Hence, our contests used a partner matching protocol (*à la* Baik et al., 2022) for studying repeated interaction situations that occur frequently in contests outside the laboratory, such as in the work-place settings.

Parts 2 and 3 featured the same real-effort task as in Part 1, but the monetary payoffs were determined by relative performance within pairs: the subject with the higher score won a prize of 285 ECU (€8.55), whereas the other received a payment of 35 ECU (€1.05). In case of a draw, the winner was determined randomly. As in Part 1, subjects played eight identical rounds.

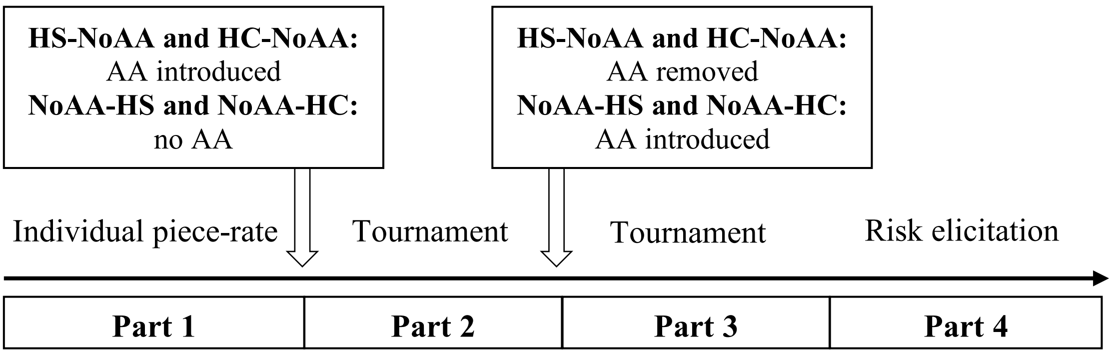
The AA policies varied between treatments: we implemented either head-start (HS) or handicap (HC) to close gaps in ability scores between the high and low-ability types. Head-start brought the low ability players (Category III) 12 extra points in each round. Conversely,

[3] A detailed look in the distribution of ability scores by gender reveals that there is a significant gender difference in the composition of types. Precisely, there are fewer women among the high ability Category I than men (15.4% of all female subjects are in Category I vs. 57.4% of all male subjects are in Type I). Similarly, the average ability score of women (8.25 correctly solved questions) is significantly lower than the one of men (11.26, $p < 0.001$, FP2S test).

handicap reduced the score of the high ability players (Category I) by 12 points in each round.[4] Obviously, the two policies differ only in terms of framing.

In two out of four treatments, the AA policy was introduced only in Part 2, and not in Part 3. In the other two treatments, it was introduced only in Part 3, and not in Part 2. Henceforth, we refer to the treatments with the AA policies in Part 2 as 'HS-NoAA' and 'HC-NoAA' (head-start and handicap, followed by an interaction in Part 3 without an AA policy), and to the treatments with AA policies in Part 3 as 'NoAA-HS' and 'NoAA-HC'. Subjects had full information about the respective AA policy when it was implemented, but they did not learn about the other AA policy and about a potential later removal of the policy. Everything else was identical between the treatments and the two parts: contestants' categories, the pair composition, the underlying task and the monetary incentives for winning or losing the contest.

FIGURE 3. EXPERIMENTAL SETUP



Notes: Each part consisted of eight identical rounds. Instructions were distributed before the first round of each part.

Our motivation for this setup is the following. Based on the difference in achievements, demographic characteristics (such as gender, race, caste, origin, etc.) are often used to determine the target group for an AA policy. For example, in the United Kingdom, the Athena Swan policy is targeted towards support of females in academia. In India the 'reservation' policy is targeted towards supporting the people of lower castes, tribal origin, and 'other backward castes' (OBCs). Similarly, in Australia, policies are directed towards indigenous people (Aboriginal and Torres Strait Islander Australians). A common feature of these policies is to, first, identify the disadvantaged group, and then to introduce AA policies to increase their chances. Of course, the reasons for why a specific group is disadvantaged and why AA seems

[4] As observed in Part 1 and in the pre-test of the experiment, low-ability types scored on average 13 points less per round than high-ability types. An AA of 12 points allowed the low ability subjects to lift up their (ex-ante) chances to win a tournament from 3% (as in parts without AA) to 22%-32% when AA policies applied.

justified are usually embedded in a historical background, and it is impossible to bring such backgrounds into experimental laboratory. However, we capture the incentives and the essence of it in our setup: first, we categorize subjects that are 'left behind' in the mathematical task, and we then support them with the AA policy (for more examples, see Chowdhury et al., 2023).

At the end of each round, subjects were informed about their score and the score of their opponent, they were reminded about the AA policy (if applicable), and they learned about the resulting bonus (reward) allocation. Subjects also saw the results from the past rounds of the respective parts in the form of a history table. Figure 3 gives an overview of experimental parts and treatments.

4.4. Sabotage

One of the main goals of our experiment was to assess how sabotage behavior differs between treatments and ability types. Following Berger et al. (2013), we implemented sabotage in a way such that every contestant could negatively influence the score of their opponent. During a tournament round everyone could click the 'block' button and blackout the screen of the opponent for nine seconds, preventing the opponent from performing the task. The consequence for the saboteur him- or herself (i.e., the cost of sabotaging) was a three-second-long blackout. Subjects could use the 'block' button as often as they desired, but at least 12 seconds had to pass in-between two consecutive clicks by the same person (i.e., nine seconds of sabotage-time and an additional three seconds of a cooling-off period). All this was common knowledge among subjects.

To reduce the possibility that subjects click on the 'block' button, because it represented the only alternative to performing the real-effort task – perhaps out of boredom – we provided an additional activity option: a 'break' button. After clicking on this button, subjects could 'take a break' and read cartoons that appeared for nine seconds (see Online Appendix III for an example). There was no limit on the number of breaks and no costs for taking them. However, subjects did not use this option much, as the average number of breaks amounts to 0.41 (i.e., less than 4 seconds) per subject per round.

4.5. Risk Preferences and Demographic Characteristics

In Part 4 we elicited subjects' risk preferences. The task was a close resemblance of the single choice list risk-elicitation task from Eckel and Grossman (2008): each subject was presented six lotteries, with two equally likely prizes for each. The expected payoffs from the lotteries increased with their increasing variances. Subjects had to choose one lottery that they would like to play. Then the computer 'tossed a coin' and determined the lottery's outcome according to the chosen option. At the end of Part 4, subjects learned about the realization of the lottery and the randomly selected rounds in Parts 1 to 3 that were payoff relevant. We also elicited self-reported risk attitudes following Dohmen et al. (2011) and demographic characteristics. Finally, subjects received their earnings privately and in cash.

5. Data and Identification Strategy

In the empirical analysis, we proceed as follows. First, we compare subjects' behavior between treatments in Part 2 across the dimensions of different AA policies. To do so, we pool the data from NoAA-HS and NoAA-HC, as all subjects faced the same conditions (no AA). Second, we proceed with a within-subject analysis of response strategies towards the introduction of AA policies by comparing Part 2 with Part 3 in the NoAA-HS and NoAA-HC treatments, separately for the two AA policies. Third, we investigate the reactions to the removal of AA policies by comparing Part 2 with Part 3 in the HS-NoAA and HC-NoAA treatments, again separately for the two AA policies. We describe results for effort provision in Section 6 and for sabotaging in Section 7.

As a measure of effort, we use the number of correctly solved problems per *available working minute*. Since the initially available working time of two minutes per round could be reduced by being sabotaged, we subtract received sabotage time (in seconds).[5] Specifically, we define effort as:

$$effort_{it} = \frac{number\ of\ correctly\ solved\ problems_{it}}{120 - received\ sabotage\ seconds_{it}} * 60$$

where i is the individual subject's identifier and t is the round's number.

We do not deduct the time lost for imposing sabotage on the opponent. This is because the application of sabotage is one's own conscious decision.

[5] We look at the number of correctly solved problems. For an alternative effort measure, we could resort to the total number of *submitted* answers instead of the number of *correct* answers. In any case, these two measures are highly correlated (Spearman's $\rho = 0.92$ and $p < 0.01$ for both types, see Figure A1 in Appendix I), and empirical results are largely the same.

Our measure for sabotage is the number of sabotage seconds per minute of round t that subject i imposed on her opponent. For easier interpretation, we normalize the measure by seconds of available working time, i.e., we deduct the received sabotage seconds from the two minutes:

$$sabotage_{it} = \frac{imposed\ sabotage\ seconds_{it}}{120 - received\ sabotage\ seconds_{it}} * 60$$

In addition, we focus on three measures: (i) the share of subjects that never use sabotage; (ii) *sabotage occurrence* as the number of rounds in which a subject committed at least one act of sabotage (i.e., number of rounds with $sabotage_{it} > 0$); (iii) *sabotage intensity* as sabotage committed by saboteurs (i.e., $sabotage_{it}|sabotage_{it} > 0$).

6. Results: Affirmative Action and Effort Provision

6.1. Types of Affirmative Action and Effort Provision

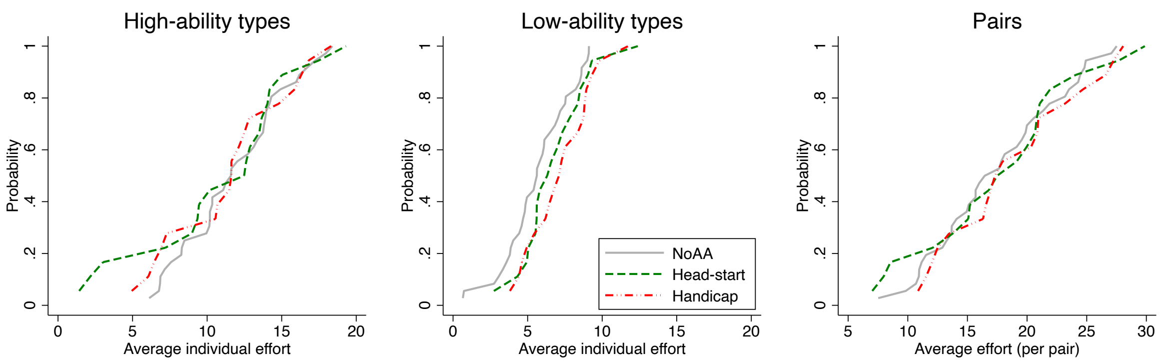
Figure 4 shows cumulative distribution functions of average individual effort levels.[6] The results of pairwise comparisons of effort under different regimes tested with FP2S tests are not significant for high-ability types (all $p > 0.45$).[7] However, the effort of low-ability types is significantly higher under the AA regimes than without AA ($p = 0.09$ for HS vs. NoAA, and $p = 0.02$ for HC vs. NoAA, FP2S test). The difference between head-start and handicap is not significant (all $p > 0.55$, FP2S test). Focusing only on the first round of Part 2 does not change these results.

We estimate a random-effects panel regression model of individual effort on a constant, the AA dummies, an ability type dummy, and control variables such as experimental round, gender, and age. As reported in Table 1, the results are largely in line with the non-parametric test results: Model (1) suggests that being 'born into' an AA policy has, on average, no effect on effort levels; coefficients of the dummy variables *Head-start* and *Handicap* are not significantly different from zero. Adding interaction terms of AA policies with the dummy *Low-ability* as explanatory variables in model (2), enables us to disentangle the impact of the AA policy on players of different types. In model (2), the coefficients *Head-start* and *Handicap*

[6] Panel (A) of Table 1 reports descriptive statistics of average individual efforts in Part 2 under the three regimes.
[7] When reporting non-parametric statistics, we generally focus on individual averages over eight rounds. This enables us to ensure statistical independence of our observations. All reported tests are two-sided.

measure the impact of the initially present AA policies on the effort of high-ability types and are not significantly different from zero. However, the statistically significant and positive coefficient of the interaction term *Handicap* × *Low-ability* indicates that handicap increases the effort levels of low-ability subjects by 0.51 standard deviations ($p = 0.02$, two-sided Wald test), as compared to a regime without AA. This result is also true for head-start if we control for lagged effort in model (3) ($p = 0.09$, Wald test). Neither non-parametric tests nor the post regression estimates provide statistical support for any differences in the impact of head-start and handicap on effort provision (all $p \geq 0.62$, Wald tests).

FIGURE 4. EMPIRICAL CDFS OF EFFORT IN PART 2



Notes: Cumulative distribution functions of average efforts over eight rounds. One independent observation per subject in the left and the middle panels. One independent observation per pair in the right panel.

Result 1. *In Part 2, when AA policies are implemented right at the beginning: (a) the AA policies do not significantly affect overall effort levels; (b) effort provision of high-ability types is not significantly different in regimes with or without AA; (c) low-ability types exert more effort when an AA policy is in place than when not; and (d) effort levels exerted under head-start and handicap are not significantly different from each other.*

Results 1(a) and 1(b) contradict Hypothesis 1, regarding increased effort under AA policies. Result 1(c) does not refute Hypothesis 1, but only for low-ability types. Result 1(d) is in line with Hypothesis 3, predicting the same effort reaction towards handicap and head-start.

TABLE 1. IMPACT OF AA POLICIES ON EFFORT IN PART 2

Independent variables	(1)	(2)	(3)
Head-start	0.080 (0.840)	-0.721 (1.272)	-0.398 (0.353)
Handicap	0.551 (0.775)	-0.392 (1.060)	-0.236 (0.325)
Low-ability	-4.608*** (0.491)	-5.514*** (0.485)	-1.212*** (0.262)
Head-start × Low-ability		1.608 (1.065)	0.793** (0.381)
Handicap × Low-ability		1.893** (0.801)	0.652** (0.306)
Effort in t-1			0.810*** (0.037)
Constant	10.863*** (1.489)	11.354*** (1.546)	2.363*** (0.596)
# Observations	1,152	1,152	1,008
# Pairs	72	72	72
Overall R^2	0.340	0.349	0.724

Post-regression Wald-tests		p-values	
Handicap = Head-start	0.625	0.822	0.717
Head-start + Head-start × Low-ability = 0		0.134	0.085
Handicap + Handicap × Low-ability = 0		0.017	0.057
Head-start × Low-ability = Handicap × Low-ability		0.818	0.755

Notes: GLS panel regression (random effects) with $effort_{it}$ as the dependent variable. The analysis includes the data from Part 2 of all four treatments. *Head-start* and *Handicap* are subject- and time-invariant dummies for the AA policies. The reference group is the regime without AA. *Low-ability* is an indicator variable for the low-ability type (with the high-ability type as the reference group). *Head-start × Low-ability* and *Handicap × Low-ability* are interaction terms. *Effort in* $t-1$ corresponds to the effort in the previous period. Not reported control variables are age, rounds and a dummy for female subjects. Controlling for lagged effort of the rival, being the tournament winner in the preceding round, or risk attitudes does not change the results. Robust standard errors (clustered on the pair' level) are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

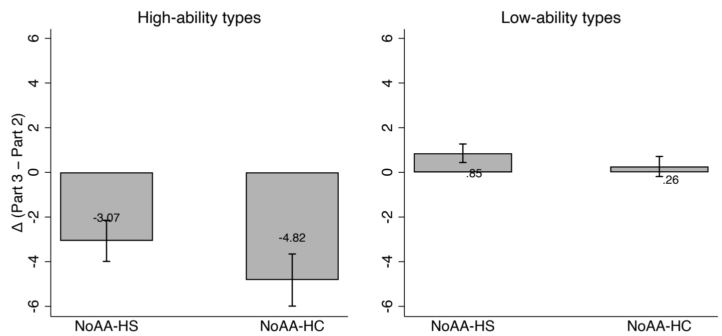
6.2. Introduction of Affirmative Action Policies and Effort Provision

Figure 5 provides an overview of changes in individual average effort levels from Part 2 to Part 3 after the introduction of the AA policies.[8] The average effort levels of high-ability types are significantly lower in Part 3 than in Part 2 (in both treatments $p < 0.01$, two-sided Fisher-Pitman permutation test for paired replicates; or FPP test, henceforth). In the NoAA-HS treatment, the average individual effort of low-ability types is weakly significantly higher in Part 3 than in Part 2 ($p = 0.06$, FPP test). However, there is no significant change in effort of low-ability types in the NoAA-HC treatment ($p = 0.58$, FPP test). In both treatments, the effort reduction of high-ability types largely outweighs the small positive change in effort levels of low-ability types. Thus, overall effort is significantly lower in Part 3 than in Part 2 ($p = 0.05$

[8] The overall pattern is very similar when we focus only on the very first rounds of Parts 2 and 3.

for NoAA-HS and $p = 0.01$ for NoAA-HC, FPP tests). Again, head-start and handicap do not differ in their effects on changes in effort levels (all $p > 0.20$, FP2S tests).

FIGURE 5. EFFORT CHANGE OF THE INTRODUCTION OF AFFIRMATIVE ACTION POLICIES



Notes: Data points are individual average efforts in two parts. Standard error bars.

TABLE 2. EFFORT AND THE INTRODUCTION OF AFFIRMATIVE ACTION

Independent variables	(1)	(2)	(3)
Introduction head-start	-1.200** (0.521)	-3.187*** (0.923)	-0.897*** (0.324)
Introduction handicap	-2.185*** (0.737)	-4.701*** (1.162)	-1.260*** (0.411)
Low-ability	-3.860*** (0.592)	-6.076*** (0.600)	-1.435*** (0.277)
Introduction head-start × Low-ability		3.999*** (0.978)	1.108*** (0.360)
Introduction handicap × Low-ability		5.007*** (0.959)	1.679*** (0.392)
Effort in t-1			0.798*** (0.035)
Constant	8.096*** (1.957)	9.259*** (1.925)	2.577*** (0.630)
# Observations	1,152	1,152	1,008
# Pairs	72	72	72
Overall R^2	0.183	0.230	0.692
Post-regression Wald-tests	p-values		
Introduction HS = Introduction HC	0.280	0.311	0.474
Introduction HS + Introduction HS × Low-ability = 0		0.048	0.216
Introduction HC + Introduction HC × Low-ability = 0		0.492	0.009
Introduction HS × Low-ability = Introduction HC × Low-ability		0.462	0.209

Notes: GLS panel regression (random effects) with $effort_{it}$ as the dependent variable. The data are from Parts 2 and 3 of the NoAA-HS and NoAA-HC treatments. *Introduction head-start* and *Introduction handicap* are dummy variables that equal one in Part 3 under the respective AA policy and zero otherwise. The reference group is Part 2 when no AA policies applied. *Low-ability* is an indicator variable for the low-ability type (with the high-ability type as the reference group). *Introduction head-start × Low-ability* and *Introduction handicap × Low-ability* are interaction terms. *Effort in* $t-1$ corresponds to the effort in the previous period. Not reported control variables are age, rounds and a dummy for female subjects. Controlling for lagged effort of the rival, being the tournament winner in the preceding round, and risk attitudes does not change the results. Robust standard errors (clustered on the pairs' level) are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2 presents random effects regression estimations of introduced AA policies on effort. In general, the regression analysis supports the non-parametric test results. Namely, in model (1), the coefficients of the dummy variables *Introduction head-start* ($p < 0.05$) and *Introduction handicap* ($p < 0.01$) indicate an overall negative effect of the introduction of the AA policy on effort levels. In models (2) and (3) we add the interaction terms between the ability type and the AA policy. The positive and at the 1% level significant coefficients show the differential effects of the AA policy on high- and low-ability types. The introduction of AA policies has a significantly negative effect on the effort levels of the high-ability types, but a significantly smaller effect on low-ability types, with no evidence for a significant negative effect of AA policies on the effort levels for low-ability types (see post-regression test estimations). Again, neither the non-parametric results nor the post-regression test of equality of the coefficients provide any evidence for different effort effects of the handicap and head-start.

Result 2. *(a) The overall effect of the introduction of AA policies on effort levels is negative. (b) Effort provision of high-ability types goes down after the introduction of the AA policies in Part 3. (c) Effort of low-ability types does not decline, after the AA policies are introduced. (d) The effects of the introduction of head-start and handicap policies on effort provision are not significantly different from each other.*

Results 2(a), 2(b) and 2(c) do not support Hypothesis 1 that predicts higher effort after the introduction of AA policies. Overall, our results strongly indicate that high-ability contestants will not increase their effort levels under AA policies if they experience competition earlier without AA. Result 2(d) provides support for Hypothesis 3 as we observe no differences in the effect of the two policies. So far, we do not find much support for positive effects of the introduction of AA policies on effort in our setup.[9]

Is it possible that the lower effort of high-ability types in Part 3 under AA than in Part 2 is due to a general fatigue or dynamic effects unrelated to AA? At least three observations suggest that this is unlikely. Firstly, there is no decrease in effort under the AA-free regime in Part 3. Indeed, effort of both types is not statistically different in Parts 2 and 3 of NoAA (all $p \geq 0.36$, FP2S tests). Table A2 in the Appendix support the non-parametric results. Secondly, we do not

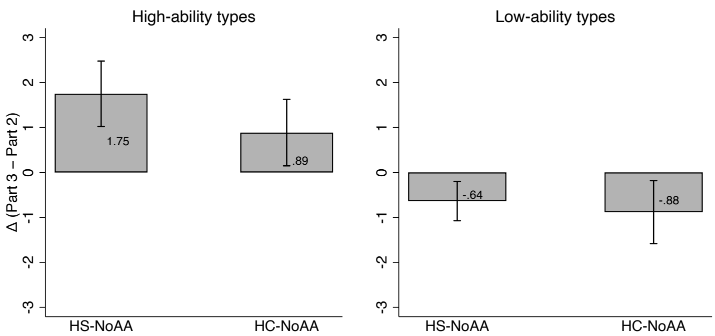
[9] In Online Appendix II, we present results for the direct comparison of subjects' behavior under the AA regimes in Part 2 and in Part 3, abstracting from uncontrolled potential effects from the history of the interaction. We find similar patterns as described in Result 1 and Result 2.

observe any indication of fatigue or similar effort-decreasing effects for low-ability types. Thirdly, as presented in the next chapter, high-ability types increase their effort provision after the removal of the AA policies, suggesting that there is no sign for general fatigue effects.

6.3. Removal of Affirmative Action and Effort Provision

Figure 6 illustrates the difference in average individual effort levels between Part 3 and Part 2, for treatments HS-NoAA and HC-NoAA in which AA policies are removed.[10] It reveals that high-ability types increase their effort provision after the removal of the AA policy, but the increase is only statistically significant for head-start (p = 0.02, FPP test), and not for handicap (p = 0.25, FPP test). Low-ability types slightly decrease effort in Part 3 when an AA policy is removed, although this decline is not statistically significant (both $p \geq 0.16$, FPP tests). The overall (across both types) change in effort provision is not significantly different from zero ($p \geq 0.16$, FPP tests). In other words, there is no reduction of total effort, as predicted by theory. Again, the difference in reaction between handicap and head-start is not statistically significant ($p \geq 0.41$, FP2S tests).

FIGURE 6. EFFORT CHANGE OF THE REMOVAL OF AFFIRMATIVE ACTION POLICIES



Notes: Data points are individual average efforts in two parts. Standard error bars.

Table 3 reports results from the econometric analysis.[11] In model (1), the coefficients of the dummy variables *Removal head-start* and *Removal handicap* measuring the average effect of the removal of AA policies on effort provision, are not significantly different from zero. Model (2) includes the interaction terms of *Low-ability* and the dummy variables for the

[10] Out-of-laboratory examples of such removals include the 1996 ban on the California State University system from recruiting and offering scholarships to students based on race, and the Equality Act 2010 in the UK that prohibits most of workplace AA (except, or instance, in the context of disability).
[11] See Table A3 in the Appendix for estimations of an alternative specification that lumps AA policies together.

removal of AA policies. While the coefficient of the *Removal head-start* dummy becomes positive at the 5%-significance level, the coefficient of its interaction term with an ability dummy is negative and significant ($p < 0.01$). Furthermore, the interaction dummy *Removal handicap* × *Low-ability* is negative and marginally significant ($p = 0.05$). The parametric results support the general picture from the non-parametric tests: while the removal of head-start has a positive effect on the effort of high-ability types, its effect on the effort of low-ability types is significantly smaller, and the overall effect is not significantly different from zero. We do not observe any significant difference in the reaction to the removal of head-start and handicap policies.

TABLE 3. EFFORT AND THE REMOVAL OF AFFIRMATIVE ACTION POLICIES

Independent variables	(1)	(2)	(3)
HS-NoAA treatment	-0.417 (0.977)	-0.418 (0.978)	-0.078 (0.288)
Removal head-start	0.556 (0.377)	1.755** (0.711)	0.512** (0.256)
Removal handicap	0.002 (0.553)	0.881 (0.727)	0.369 (0.264)
Low-ability	-4.709*** (0.631)	-3.662*** (0.651)	-0.404 (0.284)
Removal head-start × Low-ability		-2.398*** (0.903)	-0.859*** (0.333)
Removal handicap × Low-ability		-1.758* (0.905)	-0.608* (0.321)
Effort in t-1			0.814*** (0.044)
Constant	11.860*** (2.348)	11.292*** (2.356)	2.255*** (0.807)
# Observations	1,152	1,152	1,008
# Pairs	72	72	72
Overall R^2	0.326	0.338	0.733
Post-regression Wald-tests	p-values		
Removal HC = Removal HS	0.407	0.383	0.651
Removal HS + Removal HS × Low-ability = 0		0.137	0.083
Removal HC + Removal HC × Low-ability = 0		0.212	0.371
Removal HS × Low-ability = Removal HC × Low-ability		0.612	0.528

Notes: GLS panel regression (random effects) with $effort_{it}$ as the dependent variable. The data are from Parts 2 and 3 of the HS-NoAA and HC-NoAA treatments. *HS-NoAA treatment* is a treatment dummy that equals one in the treatment HS-NoAA and zero otherwise. The reference group is HC-NoAA treatment. *Removal head-start* and *Removal handicap* are dummy variables that equal one in Part 3 of the respective treatment when AA policy was removed and zero otherwise. The reference group is Part 2 when the respective AA policies applied. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Removal head-start* × *Low-ability* and *Removal handicap* × *Low-ability* are interaction of the dummy variables. *Effort in* $t-1$ corresponds to the effort in the previous period. Not reported control variables are age, rounds and a dummy for female subjects. Controlling for lagged effort of the rival, being the tournament winner in the preceding round, and risk attitudes does not change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Result 3. *(a) The overall effect of the removal of AA policies on effort is not negative. (b) Effort levels of high-ability types increase after the removal of AA policies in Part 3. (c) Compared* *to the efforts of high-ability types, effort levels of low-ability types decrease after the removal of the AA policies. (d) The adjustment of effort levels is similar for head-start and handicap.*

Results 3(a), 3(b), and 3(c) contradict Hypothesis 2, predicting decreasing effort after the removal of an AA policy. Result 3(d) is in line Hypothesis 3. It seems that, when an AA policy is removed, the removal brings about a discouragement effect for low-ability types and an encouragement effect for high-ability types. This can dramatically reduce the winning chances of low-ability types.

7. Results: Affirmative Action and Sabotage Behavior

7.1. Types of Affirmative Action and Sabotage Behavior

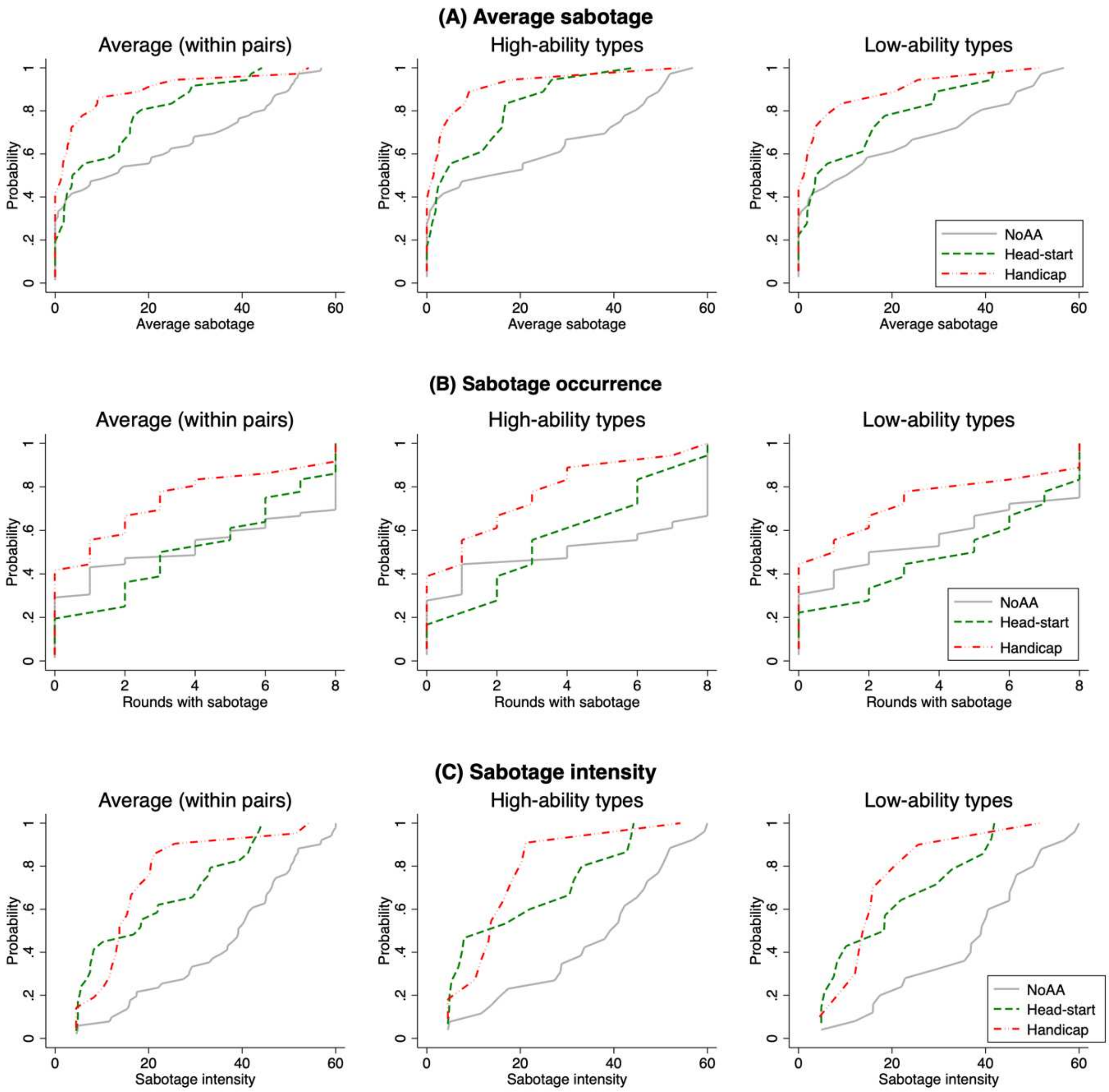
Figure 7 displays the cumulative distribution functions of average sabotage, sabotage occurrence and sabotage intensity in Part 2. Panel (A) illustrates empirical CDFs of individual average number of sabotage seconds exerted per minute over all rounds. On average, subjects sabotage for 19.6 seconds per minute when there is no AA policy in place, 11.4 under the head-start policy, and 6.4 under the handicap policy.[12] The difference in sabotage seconds between handicap and no AA is statistically significant ($p = 0.01$ for comparison of pairs, $p = 0.01$ for high-ability types and $p = 0.03$ for low-ability types, FP2S tests). Furthermore, average sabotage committed by high-ability types in Part 2 is significantly lower under head-start than without AA ($p = 0.03$, FP2S test) but not for low-ability types ($p = 0.26$, FP2S test). Table A5 provides results of the panel regression estimations for overall sabotage levels. As can be seen, the coefficient for the treatment dummy *head-start* is negative and weekly significant, and the treatment dummy for *handicap* is negative and highly significant, suggesting that average sabotage is lower under both AA policies than without an AA policy. The non-parametric tests of average sabotage provide no evidence for a significant difference between head-start and handicap (all $p \geq 0.23$, FP2S tests).

Panel (B) shows the empirical distributions of sabotage occurrence (i.e., the number of rounds when $sabotage_{it} > 0$ that ranges from 0 to 8). Under handicap, sabotage occurred in significantly fewer rounds than under head-start ($p = 0.05$) or without AA ($p = 0.07$, FP2S test of pair averages). When assessing average sabotage occurrence by type, both types sabotage in

[12] In Panel (A) of Table A4 in Appendix I, we present descriptive statistics for sabotage behavior under the three regimes.

weakly significantly fewer rounds under handicap than under head-start (p = 0.07 for both low-ability types and high-ability types, FP2S tests). However, the difference between handicap and the regime without AA is significant only for high-ability types (p = 0.03). In a similar vein, the share of subjects that never sabotage in Part 2 is highest under handicap (42%) and lowest under head-start (20%) (see also Table A4 in the Appendix for more details).

FIGURE 7. EMPIRICAL CDFS OF SABOTAGE IN PART 2



Notes: Cumulative distribution functions of individual sabotaging behavior (average over eight rounds) in Part 2. One independent observation per subject (or pair).

Panel (C) reports the distributions of sabotage intensity (when $sabotage_{it} > 0$). On average, 'actively' sabotaging contestants of both types sabotage less intensively under AA policies than without AA (all $p \leq 0.02$, FP2S tests). There is no significant difference between handicap and head-start (all $p \geq 0.58$, FP2S tests).[13]

We conduct our econometric analyses of sabotage, using a panel hurdle model (for technical details see Engel and Moffatt, 2014). This approach allows us to analyze sabotage occurrence and intensity simultaneously. Based on the logic of the hurdle model, we assume that a subject's decision to sabotage in round t and the level of sabotage follow two separate stochastic processes. In the first specification, we estimate the probability of (any) potential sabotaging behavior as a function of the AA policy, round dummies, subjects' type and demographic control variables such as gender and age by using a panel probit regression with random effects and standard errors clustered by subjects' pairs. Next, we estimate a tobit model on the restricted sample of those who sabotaged. The dependent variable in the tobit regressions is equal to $sabotage_{it}$ but the sample is restricted to data points meeting the $sabotage_{it} > 0$ condition. It is interval-truncated between 1 and 60 (seconds), with the same basic set of control variables as for the panel probit regression. In models (2) and (3), we add interaction variables of the style *[AA Policy]* × *Low-ability type*. In model (3), we include the amount of received sabotage in the previous round as an independent variable.[14]

The estimation results of probit regressions in Table 4 confirm that the head-start policy, on average, does not have any effect on the probability of sabotage occurrence as compared to the regime without AA. Handicap, however, decreases the probability of sabotage occurrence, on average, by 0.21 (see first specification of model (1), $p < 0.05$). The significant results of the post-regression tests for the difference between the coefficients of *Head-start* and *Handicap* suggest that the probability of sabotage occurrence is significantly lower under handicap than under head-start ($p < 0.05$). The interaction dummy *Handicap* × *Low-ability* in columns three and five is positive and weakly significant ($p < 0.1$), suggesting that the decrease in the probability of sabotage occurrence is more pronounced among high-ability types. Indeed, the probability of the low-ability types to apply sabotage in one round is not different under handicap than under the other two regimes (the linear combination of the coefficients *Handicap* and *Handicap* × *Low-ability* is negative, but it is not significantly different from zero,

[13] The overall patterns of these indicators are very similar when we focus on the very first round in Part 2.
[14] The standard GLS estimations were run to check robustness and are reported in Table A6 in Appendix I.

$p = 0.15$). This supports the non-parametric results: high-ability types use sabotage in fewer rounds under handicap than under other regimes.

TABLE 4. IMPACT OF AFFIRMATIVE ACTION ON SABOTAGE IN PART 2

	(1)		(2)		(3)	
Independent variables	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*
Head-start	0.035	-17.152***	-0.018	-16.169***	-0.013	-14.519***
	(0.095)	(3.651)	(0.099)	(5.039)	(0.095)	(4.142)
Handicap	-0.208**	-20.971***	-0.254***	-21.564***	-0.224**	-18.232***
	(0.097)	(4.096)	(0.095)	(5.702)	(0.088)	(4.826)
Low-ability	0.085*	2.162	0.031	2.590	-0.000	-1.008
	(0.049)	(3.491)	(0.060)	(4.677)	(0.065)	(3.820)
Head-start × Low-ability			0.108	-2.049	0.160	2.926
			(0.084)	(7.235)	(0.118)	(5.928)
Handicap × Low-ability			0.094*	1.191	0.142*	4.569
			(0.055)	(8.140)	(0.073)	(6.831)
Sabotage received in t-1					0.009***	0.590***
					(0.003)	(0.066)
Constant		31.440***		30.896***		30.997***
		(9.598)		(9.905)		(8.248)
# Observations	1152	499	1152	499	1008	445
# Pairs	72	52	72	52	72	50
Wald Chi^2	30.501	82.159	43.333	82.339	73.130	154.758
Post-regression Wald-tests	p-values					
HS = HC	0.013	0.406	0.008	0.396	0.036	0.486
HS + HS × Low-ability = 0			0.406	0.001	0.146	0.007
HC + HC × Low-ability = 0			0.149	0.001	0.443	0.006
HS × Low-ability = HC × Low-ability			0.866	0.720	0.886	0.827

Notes: The data comprise of one observation per individual per round (Part 2, all treatments). Estimation coefficients in columns one, three and five report panel probit regressions with random effects clustered on the pairs' level. The dependent variable *Sabotage occurrence*$_{it}$ is a dummy variable that takes the value 1 in case of sabotage by subject *i* in round *t*. These coefficients show marginal effects. Estimation coefficients reported in columns two, four and six report panel tobit regressions. The dependent variable here is *Sabotage intensity*$_{it}$, conditioned on $sabotage_{it} > 0$. The data points $sabotage_{it} = 0$ are not included in the tobit estimations. Therefore, the dependent variable is truncated between 1 and 60. *Head-start* and *Handicap* are subject- and time-invariant dummies for the AA policy. The reference group is the regime with no AA. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Head-start* × *Low-ability* and *Handicap* × *Low-ability* are interaction dummies. *Sabotage received in t – 1* corresponds to the sabotage received in the previous round. Not reported control variables are age, round dummies and a female dummy. Controlling for risk attitude and being the tournament winner in the preceding round does not substantially change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The tobit models show that subjects of both types sabotage less intensively under head-start and handicap as compared to when no AA policy applies (the effect size ranges between 14.5 to 21.6 (seconds of sabotage per available working minute), all $p < 0.01$). With an average sabotage intensity of 36.5 seconds under no AA, the effect size is quite substantial in magnitude (for more details see Panel (A) in Table A4 in Appendix I). There is no significant difference in this pattern between head-start and handicap, nor between high-ability and low-ability types, as can be seen from the post-regression test results and interaction variables.[15]

Result 4. *In Part 2, when AA policies apply right from the beginning: (a) the number of rounds with sabotage is significantly lower under handicap than under head-start or without an AA regime; (b) sabotage intensity of both types is significantly lower with AA policies than without; and (c) there is no significant difference between handicap and head-start with regard to sabotage intensity, but there is one for sabotage occurrence (lower under handicap than under head-start).*

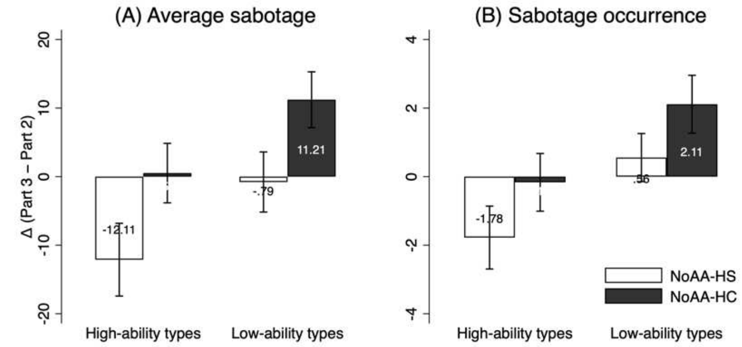
Results 4(a) and (b) clearly reject Hypothesis 1 predicting higher sabotage under the AA policies. Result 4(c) partially rejects Hypothesis 3, given that sabotage is occurring less frequently under handicap than under head-start. Since our theory does not explicitly capture behavioral motivations such as fairness preferences, social norms, reference points, emotions, etc. that may affect sabotage decisions in a particular context, the observed differences warrant further research.

7.2. Introduction of Affirmative Action and Sabotage Behavior

Panel (A) of Figure 8 illustrates the changes in average sabotage levels (over all rounds and all subject), when AA policies are introduced in treatments NoAA-HS and NoAA-HC. Panel (B) indicates the changes in sabotage occurrence. When handicap is introduced in Part 3, low-ability types tend to sabotage more as compared to the regime without AA in Part 2: the average duration of sabotage event increases by 11.2 seconds per minute ($p = 0.01$) and the number of rounds with sabotage increases by 2.1 ($p = 0.02$, FPP tests). Surprisingly, we do not observe any significantly positive effects of AA policies on high-ability types' sabotage behavior. On the contrary, they sabotage, on average, less under head-start than under the regime without AA ($p = 0.04$) and in fewer rounds ($p = 0.08$, FPP tests). The change in average sabotage is (weakly) significantly different between the NoAA-HS and NoAA-HC treatments ($p = 0.08$ for high-ability types, $p = 0.06$ for low ability types, and $p = 0.04$ for pairs, FP2S tests), but the change in sabotage occurrence is not significantly different between the two treatments ($p \geq 0.18$, FP2S test).

[15] The results of the random-effects GLS panel regressions reveal a similar pattern with regard to the overall sabotage levels and are reported in Table A5 in Appendix I.

FIGURE 8. CHANGE IN SABOTAGE AFTER THE INTRODUCTION OF AFFIRMATIVE ACTION POLICIES



Notes: The sample includes Parts 2 and 3 of the NoAA-HS and NoAA-HC treatments. Bars shown are mean individual values of differences in sabotage between Part 3 and Part 2. The units in Panel (A) are seconds per work minute, and in Panel (B) rounds. The error bars are standard errors. The results are very similar when focusing only at the first rounds of the two parts.

Table 5 reports the results of regression estimations.[16] Again, we rely on a hurdle model. However, the AA dummies *Introduction head-start* and *Introduction handicap* take the value of one if the respective AA policies apply (i.e., in Part 3), and zero otherwise, with Part 2 serving as the reference group. The probit estimation in model (1) indicates that, on average, the probability of sabotage occurrence is not affected by the introduction of AA policies. However, a more nuanced effect emerges when interaction coefficients between low-ability type and the introduced AA policy are introduced in models (2) and (3). The positive and statistically significant interaction coefficients suggest that low-ability types are by 25.6 to 28.8 percentage points more likely to engage in sabotage activity after AA introduction than high-ability types under similar conditions ($p < 0.01$). Furthermore, the interaction coefficients are even larger when controlling for the amount of sabotage received in the previous period: as shown in model (3), low-ability types are by 37.5 percentage points more likely to sabotage after the introduction of head-start and by 43.7 percentage points after the introduction of handicap, relative to high-ability types ($p < 0.01$). Conversely, sabotage occurrence among high-ability types does not increase after the introduction of AA relative to rounds without AA. In fact, the negative and statistically significant estimates for head-start indicate that high-ability types are less likely to sabotage after its introduction by 16.2 to 24.6 percentage points, as compared to the regime without AA ($p < 0.1$ and $p < 0.01$, see models (2) and (3)). Additionally, the introduction of handicap decreases the likelihood of sabotage among high-ability types by 14 percentage points when controlling for received sabotage in the previous

[16] The results of random-effects GLS panel regressions reveal a similar pattern with regard to the overall sabotage levels and are reported in Table A6 in Appendix I.

period ($p < 0.1$). Note that the linear combination of the coefficients for *Introduction handicap* and *Introduction handicap* × *Low-ability* is positive and statistically significant ($p < 0.05$), suggesting that, the introduction of the handicap induces higher sabotage occurrence among low-ability types as compared to the parts without AA.

Sabotage intensity (i.e., conditional on $sabotage_{it} > 0$) falls, on average, by 5.3 seconds per available working minute, when head-start is introduced ($p < 0.01$, see column two of model (1)). This is a substantial change of 14.4% of the initial sabotage levels without AA (see Table A4 in Appendix I for mean values). In the tobit estimation of model (2), the coefficient for *Head-start* is not significant. However, the interaction term *Head-start* × *Low-ability* shows that the average decrease in sabotage intensity after the introduction of head-start as observed in model (1) is mainly driven by low-ability types who reduce their sabotage intensity after the introduction of head-start to a larger extent than high-ability types. Also, the negative and statistically significant linear combination of coefficients *Introduction head-start* and *Introduction head-start* × *Low-ability* indicates that low-ability types sabotage less strongly after the introduction of head-start than in the regime without AA ($p < 0.01$).

The coefficient for the dummy *Introduction handicap* is statistically significant, but positive in the second stage estimation of model (1) ($p < 0.05$). Thus, overall, sabotage intensity increases, after the introduction of the handicap policy by 4.3 seconds per available working minute. The interaction term *Introduction handicap* × *Low-ability* is not statistically significant, indicating that low-ability and high-ability types increase their sabotage intensity similarly after the introduction of handicap. When controlling for sabotage received in the previous round in model (3), the coefficients of *Introduction handicap* and *Introduction handicap* × *Low-ability* are not statistically significant by itself, but their linear combination is ($p = 0.02$). This suggests that low-ability types sabotage more strongly after the introduction of handicap than in the regime without AA, while holding the received sabotage constant. Most importantly, a retaliation motivation is confirmed by the significant coefficient for receiving sabotage in the previous round.

TABLE 5. SABOTAGE AND THE INTRODUCTION OF AFFIRMATIVE ACTION

Independent variables	(1) Sabotage occurrence *Panel probit*	(1) Sabotage intensity *Panel tobit*	(2) Sabotage occurrence *Panel probit*	(2) Sabotage intensity *Panel tobit*	(3) Sabotage occurrence *Panel probit*	(3) Sabotage intensity *Panel tobit*
Introduction head-start	-0.052	-5.262***	-0.162*	-0.561	-0.246***	-4.411*
	(0.074)	(1.703)	(0.084)	(2.581)	(0.094)	(2.635)
Introduction handicap	0.104	4.279**	-0.023	6.302**	-0.140*	2.261
	(0.079)	(1.811)	(0.082)	(2.695)	(0.083)	(2.787)
Low-ability	0.124	1.112	-0.018	4.583	-0.043	0.114
	(0.096)	(4.930)	(0.084)	(5.215)	(0.083)	(4.183)
Introduction head-start × Low-ability			0.256***	-8.265**	0.375***	2.437
			(0.099)	(3.424)	(0.136)	(3.624)
Introduction handicap × Low-ability			0.288***	-3.544	0.437***	3.365
			(0.094)	(3.599)	(0.122)	(3.686)
Sabotage received in t-1					0.006*	0.574***
					(0.003)	(0.061)
Constant		36.162***		35.285***		30.753***
		(12.317)		(12.521)		(10.051)
# Observations	1152	567	1152	567	1008	496
# Pairs	36	33	36	33	36	32
Wald Chi^2	17.731	36.359	43.317	44.341	68	133.271
Post-regressions Wald-tests	p-values					
Introduction HS = Introduction HC	0.137	0.000	0.228	0.058	0.401	0.070
Introduction HS + Introduction HS × Low-ability = 0			0.322	0.000	0.207	0.418
Introduction HC + Introduction HC × Low-ability = 0			0.013	0.251	0.007	0.018
Introduction HS × Low-ability = Introduction HC × Low-ability			0.818	0.328	0.749	0.846

Notes: The data comprise of one observation per individual per round (Part 2 and 3, NoAA-HS and NoAA-HC treatments). Estimation coefficients in columns one, three and five report panel probit regressions with random effects clustered on the pairs' level. The dependent variable *Sabotage occurrence*$_{it}$ is a dummy variable that takes the value one in case of sabotage by subject i in round t. The coefficients show marginal effects. Estimation coefficients reported in columns two, four and six report panel tobit regressions. The dependent variable is *Sabotage intensity*$_{it}$, conditioned on $sabotage_{it} > 0$. The data points $sabotage_{it} = 0$ are not included in the tobit estimations. Therefore, the dependent variable is truncated between 1 and 60. *Head-start* and *Handicap* are dummy variables that equal one in Part 3 under the respective AA policy and zero otherwise. The reference group is Part 2, when no AA policies applied. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Head-start* × *Low-ability* and *Handicap* × *Low-ability* are interactions of dummies. *Sabotage received in* $t-1$ corresponds to sabotage in received in the previous round. Not reported control variables are age, rounds' dummies and a female dummy. Controlling for risk attitude and being the tournament winner in the preceding round does not substantially change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

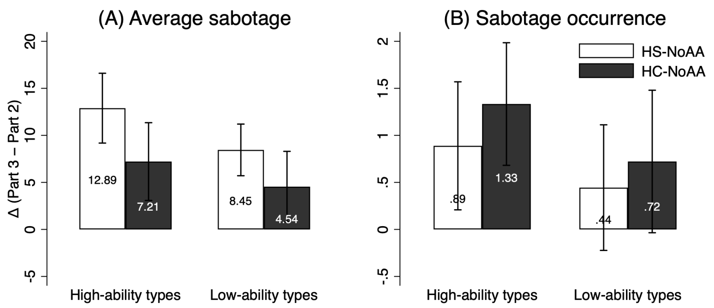
Result 5. *(a) The number of rounds in which low-ability types sabotage is significantly higher after the introduction of handicap. (b) Sabotage intensity is lower under head-start and higher under handicap than without an AA policy. (c) Handicap is less desirable, as it tends to result in more sabotaging.*

Results 5 (a) and (b) provide only partial support for Hypothesis 1. Result 5(c) contradicts Hypothesis 3. We can only speculate about potential reasons. It may be that the introduction of handicap serves as a justification to sabotage since the designer is him- or herself 'sabotaging' high-ability types with a handicap. An analogous logic could work for the introduction of head-start: the designer 'helps' low-ability types with head-start, and high-ability types react to this policy of the 'authority' by reducing sabotage. Since we provide the first experiment of this kind, these questions remain open for future research.

7.3. Removal of Affirmative Action and Sabotage Behavior

The difference between sabotage in Part 3 and Part 2 for treatments HS-NoAA and HC-NoAA, i.e., for the removal of AA policies, is illustrated in Figure 9. As can be observed in Panel (A), average sabotage is higher in Part 3 when the AA policy is removed than in Part 2. High-ability types impose, on average, significantly more sabotage after the removal of the AA policy (head-start: increase by 12.9 seconds per working minute, $p < 0.01$; handicap: by 8.5 seconds, $p = 0.09$; FPP tests). Also, low-ability types sabotage significantly more in Part 3 after the removal of head-start (head-start: by 7.2 seconds, $p < 0.01$; handicap: 4.54, $p = 0.25$; FPP tests). Panel (B) illustrates the change in the occurrence of sabotage. It suggests that the number of rounds when subjects committed sabotage is higher in Part 3. However, the change is only marginally significant in one case: for high-ability types in the HC-NoAA treatment ($p = 0.07$, FPP test). All other comparisons are not significant.

FIGURE 9. CHANGE IN SABOTAGE AFTER THE REMOVAL OF AFFIRMATIVE ACTION POLICIES



Notes: The sample includes Parts 2 and 3 of the HS-NoAA and HC-NoAA treatments. The data are based on mean individual values of differences in sabotage between Part 2 and Part 3. The error bars are standard errors.

Table 6 provides results from hurdle regression estimations for the data pooled from HS-NoAA and HC-NoAA treatments.[17] The independent variables include indicators *Removal head-start* and *Removal handicap* that are equal to one in Part 3 of the respective treatments and zero

[17] We report the separate estimations for each treatment in Table A7 in Appendix I.

otherwise. To control for Part 2 differences between the treatments HS-NoAA and HC-NoAA we also include the *HS-NoAA treatment* dummy in our estimations. The *Low-ability* dummy and its interaction with AA removal dummies are meant to capture possible differences between types. In model (3), we control for the sabotage received in the previous round. As in all previous models, we include socio-demographic control variables.

As can be seen from the (marginal effects of) probit estimates in model (1), the probability of sabotage occurrence increases, ceteris paribus, by 0.31 after the removal of head-start. The coefficient of the removal of handicap on sabotage occurrence is positive and significant in models (2) and (3) when controlling for the interaction of the removal with the ability dummy. Together with the results of the post-estimations, the probit regression estimations suggest that the high-ability types engage in sabotage more often when handicap is removed as compared to the periods when it was in place, whereas the low-ability types do not. As can be seen from the tobit estimations, sabotage intensity increases after the removal of the AA policies. The effect of the removal of head-start is especially pronounced and more than twice the size of the effect of handicap (see post-estimation results with $p < 0.05$). The interaction dummy variables of the low ability and removed AA are in almost all specifications not significantly different from zero, suggesting that both types do not differ in their reactions to the AA removal. When controlling for received sabotage (i.e., potential retaliation) in column six, we see that, again, the experience of receiving sabotage has a strong and significant effect on sabotaging the opponent.[18]

Result 6. *(a) The overall number of rounds with sabotage increases after the removal of AA policies. (b) Sabotage intensity also goes up after the removal of AA policies. (c) The effects are more pronounced after the removal of head-start than after the removal of handicap.*

Results 6 (a) and (b) contradict Hypothesis 2. Result 6(c) contradicts Hypothesis 3. Similar to the earlier results, one way to interpret this finding is that, once AA policies are removed, high-ability types may think that committing more sabotage is now admissible. At the same time, low-ability types appear not to respond with more sabotage. This is a surprising result. We

[18] The results of random-effects GLS panel regressions reveal a similar pattern regarding the overall sabotage levels and are reported in Table A7 in Appendix I.

expected that there would be more sabotage overall when a particular policy is introduced and not when it is removed.

TABLE 6. SABOTAGE AND THE REMOVAL OF AFFIRMATIVE ACTION

	(1)		(2)		(3)	
Independent variables	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*
HS-NoAA treatment	0.242^{**} (0.095)	2.136 (4.537)	0.242^{**} (0.095)	2.127 (4.531)	0.224^{**} (0.089)	16.392^{**} (6.521)
Removal head-start	0.314^{***} (0.111)	18.776^{***} (4.552)	0.342^{***} (0.127)	19.550^{***} (4.813)	0.321^{**} (0.128)	34.436^{***} (6.851)
Removal handicap	0.127 (0.079)	9.313^{***} (2.050)	0.167^{**} (0.076)	8.433^{***} (2.813)	0.127^{*} (0.075)	13.839^{***} (3.355)
Low-ability	0.119^{**} (0.050)	-2.438 (4.566)	0.153^{***} (0.057)	-2.457 (4.746)	0.173^{***} (0.066)	14.409^{**} (6.897)
Removal head-start × Low-ability			-0.056 (0.072)	-1.482 (3.036)	-0.080 (0.105)	-10.808^{***} (3.961)
Removal handicap × Low-ability			-0.079 (0.063)	1.787 (3.932)	-0.094 (0.068)	-10.693^{**} (4.676)
Sabotage received in t-1					0.008^{*} (0.004)	0.899^{***} (0.095)
Constant		15.761 (11.790)		15.349 (11.798)		4.350 (18.023)
# Observations	1152	505	1152	505	1008	446
# Pairs	72	56	72	56	72	52
Wald Chi^2	41.996	158.338	55.689	158.953	56.459	229.855
Post-regressions Wald-tests	p-values					
Removal HC = Removal HS	0.135	0.034	0.219	0.029	0.189	0.003
Removal HS + Removal HS × Low-ability = 0			0.014	0.000	0.079	0.001
Removal HC + Removal HC × Low-ability = 0			0.345	0.000	0.712	0.358

Notes: The data comprise of one observation per individual per round (Part 2 and 3, HS-NoAA and HC-NoAA treatments). Estimation coefficients in columns one, three and five report panel probit regressions with random effects clustered on the pairs' level. The dependent variable *Sabotage occurrence*$_{it}$ is a dummy variable that takes the value 1 in case of a sabotage by subject *i* in round *t*. The coefficients show marginal effects. Estimation coefficients reported in columns two, four and six report panel tobit regressions. The dependent variable is *Sabotage intensity*$_{it}$, conditioned on $sabotage_{it} > 0$. The data points $sabotage_{it} = 0$ are not included in the tobit estimations. Therefore, the dependent variable is truncated between 1 and 60. *HS-NoAA treatment* is a treatment dummy that equals one in the treatment HS-NoAA and 0 otherwise. The reference group is HC-NoAA treatment. *Removal head-start* and *Removal handicap* are dummy variables that equal one in Part 3 of the respective treatment when AA policy was removed and zero otherwise. The reference group is Part 2 when the respective AA policies applied. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Removal head-start* × *Low-ability* and *Removal handicap* × *Low-ability* are interaction of the dummy variables. *Sabotage received in t-1* corresponds to sabotage received in the previous round. Not reported control variables are age, rounds and a dummy for females. Controlling for risk attitudes and being the tournament winner in the preceding round does not change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

8. Discussion and Conclusion

AA policies are common, often controversial, and well-researched in economics, in general. However, hitherto no study has fully analyzed the effects of the lifecycle (introduction, continuation, and removal) of such policies in a controlled environment. Furthermore, comparative analyses of different AA instruments, such as head-start and handicap, have been lacking so far. Importantly, especially in organizations and workplaces, but not only there, such AA policies may affect not only effort provision, but also the level of sabotaging others. That is why our setup focuses also on sabotage.

In this study, we assess the combination of these three inter-related but erstwhile unexplored research topics. We run a laboratory experiment with a real effort tournament and the possibility of sabotaging others. We also investigate the effects of introducing and removing head-start and handicap on effort and sabotage in such a tournament.

Contrary to some earlier studies, we find that AA policies do not universally result in higher effort. In particular, high performers that already experienced contests without AA, reduce their effort after the introduction of the policy. Moreover, and reassuringly for policymakers, we observe less sabotage under AA policies, when the tournament started right away with an AA policy in place. When an existing AA policy is removed, it does not necessarily reduce effort, but unfortunately, it can significantly intensify sabotage in the short run.

These results are of importance both from the perspective of advancing scholarly literature on tournaments and from the perspective of policymaking. We first discuss some conclusions related to policymaking. Since it is almost impossible to get data allowing for causal inference on the whole lifecycle of an AA policy from outside the laboratory, policymakers often need to rely on theoretical predictions on the implications of the introduction or removal of an AA policy. Even the existing experimental studies investigate such effects in a between-subject manner. While their results can provide ample evidence on the difference in behavior between an AA policy and a regime without AA, they cannot capture the relevant dynamics of the introduction and removal of the policy itself.

In terms of the contribution to the relevant scholarly literature, our study provides a series of new techniques and results. First, we use a real-effort task and actual (not induced) individual ability to pair and match subjects in the tournament. The two AA policies that are incentive-identical but differ in the framing are implemented in a between-subject design. Our experimental setup with introducing and removing the policies is novel. Moreover, our experimental design allows comparing the effects of AA policies both between- and within-subjects as well as investigating the intertemporal effects of the experience of such a policy in another new regime without AA guiding the tournament. Finally, our design allows us to assess the extensive and intensive margins of sabotage under different policies.

Our results support some of the existing observations and add new results to the literature. The most important finding, in a nutshell, is that the lifecycle of an AA policy is very relevant for its effects on effort provision and sabotage. Therefore, only one snapshot comparing an AA policy with a baseline cannot provide the broader picture that is – given our results – necessary for an empirical assessment that informs policymakers. For example, the first experimental study on AA by Schotter and Weigelt (1992) found that an AA policy does not affect effort provision. The finding was later corroborated in a field study by Calsamiglia et al (2013). We show in this study that the conclusion may depend on the type of AA policy, the type of the player, and the chronology of the AA lifecycle, i.e., the experience. High performers that have previously experienced a regime without AA may reduce their effort provision. Moreover, both Brown and Chowdhury (2017) in the field and Leibbrandt et al. (2018) in the laboratory found that an AA policy leads to higher levels of sabotage than in an environment without AA. The current study shows that such a negative effect of an AA policy may not necessarily be universal. Contestants may even commit fewer acts of sabotage when an AA policy is introduced (as for example in case of head-start). At the same time, contestants surprisingly commit more sabotage when they already have prior experience in a regime with an AA policy, after this policy is removed.

Our results also show that head-start and handicap seem to have quite similar effects on subjects' behavior in several dimensions. Nevertheless, some differences emerge: We observe more rounds with sabotage and higher sabotage intensity imposed by low-ability types after the introduction of handicap, compared to head-start. Furthermore, the removal of head-start seems to lead to more intense sabotage than the removal of handicap.

One, however, must take caution before generalizing our results. Similar to any laboratory experiment, this study has to face the test of external validity. We introduce a very stylized real effort tournament with sabotage that may be more appropriate for organizational and workplace-related tournaments than other contests. In search of more experimental control and simplicity, we implement a two-player tournament. The number of involved players may well be relevant for the effects that we describe in the paper.

This brings us to possible extensions of our setup. Konrad (2000) showed that, when there are multiple players in a contest, committing sabotage becomes a public good, and due to the free-riding incentives in public goods, sabotage can vanish. Although field studies and laboratory experiments have shown that people still commit sabotage in multi-player settings, it is an open empirical question whether including more contestants in our design would bring different results. There is also a strand of literature that documents the effects of AA on sabotage towards the tournament designer. Ku and Salmon (2012) observed such negative effects. They randomly split subjects into high and low wage players. When the low wage players knew that they were discriminated against, they ended up producing much less in a piece-rate task. Similarly, Fallucchi and Quercia (2018) found that, when a proportion of a reward is saved for a specific group of subjects, then 'retaliation' against the experimenter increases. Our current design does not have such features, but understanding such behavior throughout the lifecycle of an AA policy seems relevant from a methodological perspective. Outside the laboratory, players often communicate with each other while taking part in tournaments. The effects of such communication on contestants' behavior when an AA policy is introduced or removed is still an open question. Investigating these questions or setups with our experimental paradigm may provide additional explanations for the behavior that we have observed in our study.

References

Amegashie, J. A., & Runkel, M. (2007). Sabotaging Potential Rivals. *Social Choice and Welfare*, 28, 143-162.

Baik, K.H., Chowdhury, S.M., & Ramalingam, A. (2022). Group Size and Matching Protocol in Contests. *Canadian Journal of Economics*, 54(4), 1716-1736.

Balafoutas, L., & Sutter, M. (2012). Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory. *Science*, 335 (6068), 579-582.

Balafoutas, L., Chowdhury, S. M., & Plessner, H. (2019). Applications of Sports Data to Study Decision Making. *Journal of Economic Psychology*, 75, 102153.

Berger, J., Harbring, C., & Sliwka, D. (2013). Performance Appraisals and the Impact of Forced Distribution - An Experimental Investigation. *Management Science*, 59(1), 54-68.

Brown, A., & Chowdhury, S. M. (2017). The Hidden Perils of Affirmative Action: Sabotage in Handicap Contests. *Journal of Economic Behavior and Organization*, 133, 273-284.

Calsamiglia, C., Franke, J., & Rey-Biel, P. (2013). The Incentive Effects of Affirmative Action in a Real-effort Tournament. *Journal of Public Economics*, 98, 15-31.

Carpenter, J., Matthews, P. H., & Schirm, J. (2010). Tournaments and Office Politics: Evidence from a Real Effort Experiment. *American Economic Review,* 100 (1), 504-517.

Che, Y.-K., & Gale, I. (2003). Optimal Design of Research Contests. *American Economic Review,* 93 (3), 646-671.

Chen, K-P. (2003) Sabotage in Promotion Tournaments. *The Journal of Law, Economics and Organization*, 19 (1), 119-140.

Chowdhury, S. M., & Crede, C. J. (2020). Post-cartel Tacit Collusion: Determinants, Consequences, and Prevention. *International Journal of Industrial Organization*, *70*, 102590.

Chowdhury, S. M., & Gürtler, O. (2015). Sabotage in Contests: A Survey. *Public Choice*, 164(1), 135-155.

Chowdhury, S. M., & Kim, S. H. (2017). "Small, Yet Beautiful": Reconsidering the Optimal Design of Multi-winner Contests. *Games and Economic Behavior*, 104, 486-493.

Chowdhury, S. M., Esteve-González, P., & Mukherjee, A. (2023). Heterogeneity, Leveling the Playing Field, and Affirmative Action in Contests. Southern Economic Journal, 89(3), 924-974.

Corchón, L. C., & Serena, M. (2018). Contest Theory. In *Handbook of Game Theory and Industrial Organization, Volume II*. Edward Elgar Publishing. Cheltenham.

Czibor, E., & Martinez, S. D. (2019). Never too Late: Gender Quotas in the Final Round of a Multistage Tournament, *The Journal of Law, Economics and Organization*, 35 (2), 319-363.

Dahm, M., & Esteve-González, P. (2018). Affirmative Action through Extra Prizes. *Journal of Economic Behavior and Organization*, 153, 123-142.

Danilov, A., Harbring, C., & Irlenbusch, B. (2019). Helping under a Combination of Team and Tournament Incentives. *Journal of Economic Behavior and Organization*, 162, 120-135.

Dato, S., & Nieken, P. (2019). Gender Differences in Sabotage: The Role of Uncertainty and Beliefs. *Experimental Economics*, 23, 353-391.

del Corral, J., Prieto-Rodríguez, J., & Simmons, R. (2010). The Effect of Incentives on Sabotage: The Case of Spanish Football. *Journal of Sports Economics*, 11 (3), 243-260.

Deutscher, C., Frick, B., Gürtler, O., & Prinz, J. (2013). Sabotage in Tournaments with Heterogeneous Contestants: Empirical Evidence from the Soccer Pitch. *Scandinavian Journal of Economics*, 115, 1138-1157.

Dohmen, T., & Falk, A. (2011). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review*, 101 (2), 556-590.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9 (3), 522-550.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices. *Journal of Economic Behavior and Organization*, 68 (1), 1-17.

Engel, C., & Moffatt, P. G. (2014). Dhreg, xtdhreg, and bootdhreg: Commands to Implement Double-hurdle Regression. *Stata Journal*, *14*(4), 778-797.

Fain, J.R. (2009). Affirmative Action Can Increase Effort. *Journal of Labor Research*, 30**,** 168-175.

Fallucchi, F., & Quercia, S. (2018). Affirmative Action and Retaliation in Experimental Contests. *Journal of Economic Behavior and Organization*, 156, 23-40.

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

Fort, R., & Maxcy, J. (2003). Competitive Balance in Sports Leagues: An Introduction. *Journal of Sports Economics*, 4(2), 154-160.

Franke, J. (2012). Affirmative Action in Contest Games. *European Journal of Political Economy*, 28(1), 105-118.

Franke, J., Kanzow, C., Leininger, W., & Schwartz, A. (2013). Effort Maximization in Asymmetric Contest Games with Heterogeneous Contestants. *Economic Theory*, 52(2), 589-630.

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

Fu, Q. (2006). A Theory of Affirmative Action in College Admissions. *Economic Inquiry*, 44(3), 420-428.

Fu, Q., & Lu, J. (2009). The Beauty of "Bigness": On Optimal Design of Multi-winner Contests. *Games and Economic Behavior*, 66(1), 146-161.

Garicano, L., & Palacios-Huerta, I. I. (2014). Making the Beautiful Game a Bit Less Beautiful. In *Beautiful Game Theory*, Princeton University Press, 124-150. Princeton.

Gill, D., & Stone, R. (2010). Fairness and Desert in Tournaments. *Games and Economic Behavior*, 69(2), 346-364.

Girard, V. (2018). Don't Touch My Road. Evidence from India on Affirmative Action and Everyday Discrimination. *World Development*, 103, 1-13.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1, 114-125.

Gürtler, O. (2008). On Sabotage in Collective Tournaments. *Journal of Mathematical Economics*, 44 (3-4), 383-393.

Gürtler, O., & Münster, J. (2010). Sabotage in Dynamic Tournaments. *Journal of Mathematical Economics*, 46 (2), 179-190.

Gürtler, O., & Münster, J. (2013). Rational Self-sabotage. *Mathematical Social Sciences*, 65 (1), 1-4.

Harbring, C., & Irlenbusch, B. (2005). Incentives in Tournaments with Endogenous Prize Selection. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift Für Die Gesamte Staatswissenschaft,* 161 (4), 636-663.

Harbring, C., & Irlenbusch, B. (2008). How Many Winners are Good to Have?: On Tournaments with Sabotage. *Journal of Economic Behavior and Organization*, 65 (3-4), 682-702.

Harbring, C., & Irlenbusch, B. (2011). Sabotage in Tournaments: Evidence from a Laboratory Experiment. *Management Science*, 57 (4), 611-627.

Harbring, C., Irlenbusch, B., Kräkel, M., & Selten, R. (2007) Sabotage in Corporate Contests – An Experimental Analysis. *International Journal of the Economics of Business*, 14 (3), 367-392.

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

Kirkegaard, R. (2012). Favoritism in Asymmetric Contests: Head Starts and Handicaps. *Games and Economic Behavior*, 76 (1), 226-248.

Konrad, K. A. (2000). Sabotage in Rent-seeking Contests. *The Journal of Law, Economics, and Organization*, 16 (1), 155-165.

Konrad, K. A. (2009). Strategy and Dynamics in Contests. *OUP Catalogue*. New York.

Kräkel, M. (2005). Helping and Sabotaging in Tournaments. *International Game Theory Review*, 7 (2), 211-228.

Krishna, K., & Tarasov, A. (2016). Affirmative Action: One Size Does Not Fit All. *American Economic Journal: Microeconomics*, 8(2), 215-52.

Ku, H., & Salmon, T. C. (2012). The Incentive effects of Inequality: An Experimental Investigation. *Southern Economic Journal*, *79*(1), 46-70.

Lazear, E. P. (1989). Pay Equality and Industrial Politics. *Journal of Political Economy*, 97 (3), 561-580.

Lazear, E., & Rosen, S. (1982). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89, 841-864.

Lee, S. H. (2013). The Incentive Effect of a Handicap. *Economics Letters*, 118(1), 42-45.

Leibbrandt, A., Wang, L. C., & Foo, C. (2018). Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash against Women. *Management Science*, 64(8), 3501-3516.

Martin, A., Mikołajczak, G., & Orr, R. (202). Does Process Matter? Experimental Evidence on the Effect of Procedural Fairness on Citizens' Evaluations of Policy Outcomes. *International Political Science Review*, 43(1), 103-117.

Moldovanu, B., & Sela, A. (2001). The Optimal Allocation of Prizes in Contests. *American Economic Review*, 91(3), 542-558.

Moldovanu, B., & Sela, A. (2006). Contest Architecture. *Journal of Economic Theory*, 126(1), 70-96.

Münster, J. (2007). Selection Tournaments, Sabotage, and Participation. *Journal of Economics and Management Strategy*, 16 (4), 943-970.

Petters, L. M., & Schröder, M. (2020). Negative Side Effects of Affirmative Action: How Quotas Lead to Distortions in Performance Evaluation. *European Economic Review*, 130, 103500.

Piest, S., & Schreck, P. (2021). Contests and Unethical Behavior in Organizations: A Review and Synthesis of the Empirical Literature. *Management Review Quarterly*, 71(4), 679-721.

Schotter, A., & Weigelt, K. (1992). Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results. *The Quarterly Journal of Economics*, 107 (2), 511-539.

Soubeyran, R. (2009). Does a Disadvantaged Candidate Choose an Extremist Position? *Annals of Economics and Statistics,* 93/94, 327-348.

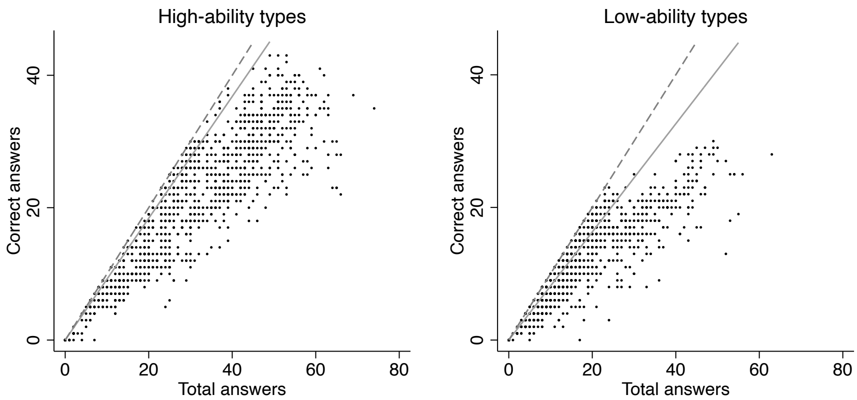
Steinmayr, A., Stracke, R., & Zegners, D. (2018). ‘Having the Lead’ vs. ‘Lagging behind’: The Incentive Effect of Handicaps in Tournaments. *SSRN Electronic Journal*.

Tasoff, J., & Zhang, W. (2022). The Performance of Time-Preference and Risk-Preference Measures in Surveys. *Management Science*, *68*(2), 1149-1173.

APPENDIX (for online publication)

Appendix I: Figure and Tables

FIGURE A1. SCATTER PLOT OF TOTAL NUMBER OF ANSWERS AND CORRECT NUMBER OF ANSWERS



Notes: Scatter plot of correct and total answers per round (of two minutes) in Parts 2 and 3. The solid line illustrates Spearman's ρ. The dashed line is the 45°-line.

TABLE A1. ABILITY SCORE BY TYPE

		Average number of correctly solved questions (per minute) in Part 1					
		Rounds 1-4			Rounds 5-8		
Type	N	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
High-ability	72	10.44	10.13	2.65	13.28	13	2.54
Mid-ability	48	7.52	7.31	1.45	9.54	9.75	0.98
Low-ability	72	5.17	5.25	1.50	6.62	6.75	1.56
Total	192	7.73	7.38	3.04	9.85	9.63	3.45

TABLE A2. DESCRIPTIVE STATISTICS OF EFFORT

Affirmative action regime	Per type or per pair	Panel (A): Part 2					Panel (B): Part 3				
		Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
No affirmative action	High-ability	11.82	11.61	3.44	6.12	18.4	12.54	12.31	3.24	6.9	20.25
	Low-ability	5.67	5.62	2.2	0.66	9.11	6.25	6.64	3.29	0	12.93
	Total	17.5	17.05	5.33	7.53	27.52	18.8	17.77	4.95	10.59	30.25
Head-start	High-ability	10.94	12.55	5.01	1.42	19.31	7.38	7.29	5.55	0	16.81
	Low-ability	6.79	6.45	2.23	2.74	12.39	6	5.91	2.46	0	12.04
	Total	17.73	18.31	6.4	7.02	29.88	13.38	12.98	7.35	2.2	28.85
Handicap	High-ability	11.51	11.63	3.97	4.94	18.31	8.38	8.58	6.29	0.57	17.88
	Low-ability	7.24	7.18	2.13	3.81	11.75	6.46	6.75	2.89	0	12.41
	Total	18.75	17.63	5.58	10.86	28.06	14.84	15.13	8.54	0.78	28.64

Notes: Medians, standard deviations, minima and maxima reported in the rows *High-ability* and *Low-ability* are computed based on the individual averages over eight rounds (i.e., with one independent observation per subject). Statistics reported in the rows *Total* represent effort levels exerted by both subjects in one pair. Here, we have one independent observation per pair. The subjects facing no AA in Part 2 were exposed to the head-start or handicap regimes in Part 3. Whereas subjects facing either head-start or handicap in Part 2 experienced no AA in Part 3.

TABLE A3. EFFORT AND THE REMOVAL OF AFFIRMATIVE ACTION POLICIES (ALTERNATIVE SPECIFICATION)

Independent variables	*(1)*	*(2)*	*(3)*
AA removed	0.279	1.041	0.431
	(0.338)	(0.652)	(0.264)
Low-ability	-4.709***	-3.670***	-0.405
	(0.631)	(0.645)	(0.282)
HS-NoAA treatment	-0.139	-0.417	-0.077
	(0.835)	(0.978)	(0.288)
Low-ability × AA removed		-2.078***	-0.733***
		(0.648)	(0.262)
HS-NoAA treatment × AA removed		0.554	0.017
		(0.669)	(0.265)
Effort in t-1			0.814***
			(0.044)
Constant	11.721***	11.340***	2.153**
	(2.305)	(2.358)	(0.886)
# Observations	1,152	1,152	1,008
# Pairs	72	72	72
Overall R^2	0.3247	0.3374	0.7331

Notes: GLS panel regression (random effects) with $effort_{it}$ as the dependent variable. The data comprise of one observation per individual per round (Part 2 and 3, HS-NoAA and HC-NoAA treatments). *AA removed* is equal one in rounds of Part 3, when no AA was in place anymore; and zero otherwise. *HS-NoAA* is a dummy taking value of one for the HS-NoAA treatment; and is zero otherwise. *Low-ability* is an indicator for the low-ability type (with high-ability types as the reference group). *Low-ability × AA removed* is an interaction of two dummy variables. *Effort in t-1* corresponds to sabotage received in the previous round. Not reported control variables are age, rounds and a dummy for females. Controlling for risk attitudes and being the tournament winner in the preceding round does not change the results. Robust standard errors (clustered on pairs level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A4. DESCRIPTIVE STATISTICS ON SABOTAGE

		Panel (A): Part 2				Panel (B): Part 3			
		% Contestants that never used sabotage	Sabotage occurrence (# rounds with $sabotage_{it} > 0$)	Intensive sabotage (conditional on $sabotage_{it} > 0$)	Average sabotage	% Contestants that never used sabotage	Sabotage occurrence (# rounds with $sabotage_{it} > 0$)	Intensive sabotage (conditional on $sabotage_{it} > 0$)	Average sabotage
NoAA	High-ability	27.78 (45.43)	4.11 (3.58)	36.10 (16.58)	20.81 (20.82)	36.11 (48.71)	3.94 (3.71)	30.96 (18.09)	18.21 (21.34)
	Low-ability	30.6 (46.72)	3.58 (3.33)	36.80 (15.47)	18.37 (19.76)	33.33 (47.81)	3.92 (3.43)	28.50 (18.48)	16.15 (19.18)
	Average (within pairs)	29.17 (45.77)	3.85 (3.44)	36.45 (15.89)	19.59 (20.19)	34.72 (47.94)	3.93 (3.55)	29.70 (18.33)	17.18 (20.17)
Head-start	High-ability	16.67 (38.35)	3.67 (2.70)	20.41 (15.77)	10.4 (12.19)	16.67 (38.35)	3.06 (2.48)	32.99 (19.20)	13.60 (15.98)
	Low-ability	22.22 (42.78)	4.33 (3.11)	20.33 (14.21)	12.43 (14.23)	11.11 (32.34)	4.67 (3.48)	30.85 (16.93)	21.49 (20.45)
	Average (within pairs)	19.44 (40.14)	4 (2.89)	20.37 (14.77)	11.42 (13.10)	13.89 (35.07)	3.86 (3.09)	31.89 (17.79)	17.55 (18.53)
Handicap	High-ability	38.89 (50.16)	2 (2.45)	17.08 (13.52)	5.91 (12.9)	11.11 (32.33)	3.22 (2.69)	35.98 (19.23)	16.41 (16.94)
	Low-ability	44.44 (51.13)	2.33 (3.05)	18.02 (13.14)	6.87 (13.33)	22.22 (42.78)	5.17 (3.59)	37.55 (15.73)	25.69 (21.43)
	Average (within pairs)	41.67 (50)	2.17 (3.21)	17.53 (18.14)	6.39 (12.94)	16.67 (37.8)	4.19 (3.28)	36.71 (17.40)	21.05 (19.61)

Notes: One independent observation per subject (in rows *Total*, per pair). Standard deviations are reported in parentheses based on independent observations. The subjects facing no AA in Part 2 were exposed to the head-start or handicap regimes in Part 3. Whereas subjects facing either head-start or handicap in Part 2 experienced no AA in Part 3.

TABLE A5. GLS REGRESSION OF IMPACT OF AFFIRMATIVE ACTION ON SABOTAGE IN PART 2

Independent variables	(1) Sabotage	(2) Sabotage	(3) Sabotage
Head-start	-7.159*	-8.898*	-9.723**
	(4.270)	(4.568)	(3.801)
Handicap	-13.507***	-15.094***	-12.016***
	(4.195)	(4.445)	(3.477)
Low-ability	3.091	1.343	-1.503
	(1.963)	(2.392)	(2.525)
Head-start × Low-ability		3.498	7.387
		(3.350)	(5.551)
Handicap × Low-ability		3.195	5.852*
		(2.012)	(3.110)
Sabotage received in t-1			0.648***
			(0.109)
Constant		29.649***	25.880***
		(8.879)	(7.151)
# Observations	1152	1152	1008
# Pairs	72	72	72
R^2	0.112	0.114	0.442
Post-regressions Wald-tests		p-values	
HS = HC	0.096	0.115	0.544
HS + HS× Low-ability = 0		0.242	0.537
HC + HC × Low-ability = 0		0.005	0.042
HS × Low-ability = HC × Low-ability		0.927	0.784

Notes: GLS panel regression (random effects) with $sabotage_{it}$ as the dependent variable. The data are from Part 2 of all four treatments. *Head-start* and *Handicap* are subject- and time-invariant dummies for the AA policy. The reference group is the regime without AA. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Head-start × Low-ability* and *Handicap × Low-ability* are the interaction of dummy variables. *Sabotage received in t-1* corresponds to the one-round lagged sabotage variable. Not reported control variables are age, rounds and a dummy for females. Controlling for lagged received or committed sabotage, or being the tournament winner in the preceding round, and risk attitudes does not change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A6. GLS REGRESSION OF SABOTAGE AND THE INTRODUCTION OF AFFIRMATIVE ACTION

	(1) Sabotage	(2) Sabotage	(3) Sabotage
Introduction head-start	-5.962	-11.552**	-15.476***
	(4.254)	(5.070)	(5.162)
Introduction handicap	5.376	-0.048	-8.214**
	(3.562)	(4.181)	(3.990)
Low-ability	5.276	-0.240	-2.808
	(3.806)	(3.467)	(3.003)
Introduction head-start × Low-ability		11.142***	18.570***
		(3.837)	(5.988)
Introduction handicap × Low-ability		10.885***	19.727***
		(3.982)	(5.806)
Sabotage received in t-1			0.564***
			(0.147)
Constant		26.399***	22.275***
		(10.127)	(8.502)
# Observations	1152	1152	1008
# Pairs	36	36	36
R^2	0.014	0.026	0.304
Post-regressions Wald-tests	p-values		
Introduction HS = Introduction HC	0.035	0.069	0.258
Introduction HS+ Introduction HS× Low-ability = 0		0.923	0.438
Introduction HC+ Introduction HC× Low-ability = 0		0.007	0.001
Introduction HS× Low-ability = Introduction HC × Low-ability		0.962	0.896

Notes: GLS panel regression (random effects) with $sabotage_{it}$ as the dependent variable. The data are from the NoAA-HS and NoAA-HC treatments. *Introduction head-start* and *Introduction handicap* are dummy variables that equal one in Part 3 under the respective AA policy and zero otherwise. The reference group is Part 2, when no AA policies applied. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Introduction head-start × Low-ability* and *Introduction handicap × Low-ability* are the interaction of dummy variables. *Sabotage received in t-1*corresponds to the one-round lagged sabotage variable. Not reported control variables are age, rounds and a dummy for females. Controlling for lagged received or committed sabotage, or being the tournament winner in the preceding round, and risk attitudes does not change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A7. SABOTAGE AND THE REMOVAL OF AFFIRMATIVE ACTION BY TREATMENT

	Panel (A): HS-NoAA treatment						Panel (B): HC-NoAA treatment					
	(1)		(2)		(3)							
Independent variables	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*
AA removed	0.076	16.618***	0.106	17.531***	0.101	14.246***	0.120	10.189***	0.158**	9.631***	0.094	7.321***
	(0.066)	(1.485)	(0.081)	(2.158)	(0.087)	(2.385)	(0.075)	(2.132)	(0.073)	(2.969)	(0.066)	(2.700)
Low-ability	0.079	-4.757	0.106	-3.928	0.109	-0.425	0.165*	0.255	0.206**	-0.369	0.133	1.573
	(0.072)	(6.179)	(0.089)	(6.313)	(0.097)	(5.055)	(0.089)	(6.918)	(0.090)	(7.319)	(0.096)	(3.625)
Low-ability × AA removed			-0.058	-1.729	-0.047	-5.359*			-0.075	1.135	-0.104	-10.934***
			(0.081)	(2.964)	(0.096)	(3.190)			(0.061)	(4.206)	(0.077)	(3.895)
Sabotage received in t-1					0.002	0.565***					0.018***	1.114***
					(0.005)	(0.089)					(0.003)	(0.098)
Constant		7.832		7.367		12.311		47.748**		47.781**		36.473***
		(14.880)		(14.834)		(11.512)		(22.252)		(22.318)		(10.980)
# Observations	576	312	576	312	504	277	576.000	193.000	576.000	193.000	504.000	169.000
# Pairs	36	30	36	30	36	30	36.000	26.000	36.000	26.000	36.000	22.000
Wald Chi^2	26.919	140.910	50.972	141.360	22.075	161.798	83.979	35.251	88.546	35.334	105.228	166.975
Post-regressions Wald-tests	p-values											
AA removed + Low-ability × AA removed = 0			0.537	0.000	0.578	0.000			0.356	0.000	0.890	0.214

Notes: Sabotage occurrence is analyzed with panel probit regressions with random effects. Marginal effects are reported. Sabotage intensity is analyzed with panel tobit regressions and conditioned on $sabotage_{it}>0$ and censored below and above at 1 and 60. The data are from Parts 2 and 3 of the HS-NoAA and HC-NoAA treatments. *AA removed* is equal to one for Part 3. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Low-ability × AA removed* is an interaction. *Sabotage received in t-1* corresponds to sabotage received in the previous round. Not reported control variables are age, rounds and a dummy for females. Controlling for risk attitudes, and being the tournament winner in the preceding round does not change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE A8. GLS REGRESSION ON SABOTAGE AND THE REMOVAL OF AFFIRMATIVE ACTION

	(1) Sabotage	(2) Sabotage	(3) Sabotage
HS-NoAA treatment	5.848	5.844	2.399
	(3.818)	(3.817)	(2.666)
Removal head-start	16.521***	18.724***	13.489**
	(5.844)	(6.591)	(5.933)
Removal handicap	5.878	7.222*	5.640*
	(3.853)	(4.097)	(3.283)
Low-ability	1.671	3.469	4.590
	(1.866)	(2.278)	(2.936)
Removal head-start × Low-ability		-4.415	-6.928
		(2.792)	(4.790)
Removal handicap × Low-ability		-2.689**	-4.783*
		(1.362)	(2.540)
Sabotage received in t-1			0.752***
			(0.198)
Constant	12.147	11.128	11.760*
	(11.469)	(11.447)	(6.911)
# Observations	1,152	1,152	1,008
# Pairs	72	72	72
R^2	0.097	0.103	0.492
Post-regressions Wald-tests	p-values		
Removal HC = Removal HS	0.103	0.120	0.195
Removal HS + Removal HS × Low-ability = 0		0.008	0.118
Removal HC + Removal HC × Low-ability = 0		0.224	0.741

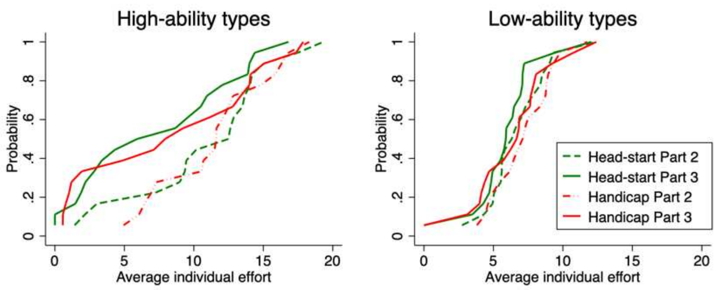
Notes: GLS panel regression (random effects) with *sabotage*$_{it}$ as the dependent variable. The data are from the HS-NoAA and HC-NoAA treatments. *HS-NoAA treatment* is a treatment dummy that equals one in the treatment HS-NoAA and 0 otherwise. The reference group is HC-NoAA treatment. *Removal head-start* and *Removal handicap* are dummy variables that equal one in Part 3 of the respective treatment when AA policy was removed and zero otherwise. The reference group is Part 2 when the respective AA policies applied. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Removal head-start* × *Low-ability* and *Removal handicap* × *Low-ability* are interaction of the dummy variables. *Sabotage received in t-1* corresponds to the one-round lagged effort variable. Not reported control variables are age, rounds and a dummy for females. Controlling for the lagged received or committed sabotage, or being the tournament winner in the preceding round, and risk attitudes does not change the results. Robust standard errors (clustered on the pairs' level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix II: Additional results

AII.1. Experience of Existing Affirmative Action and Effort Provision

As shown in 6.2., the effects of AA policies are less positive when subjects have previously entered the tournament without AA policy. In what follows, we directly compare subjects' effort levels under the AA polices in Part 2 with effort under the same conditions in Part 3. In order to do so, we pool the data from Part 2 of the HS-NoAA and HC-NoAA treatments and Part 3 of the NoAA-HS and NoAA-HC treatments. Figure AII.1 illustrates the empirical distributions of average individual effort by subject's types.

FIGURE AII.1. EMPIRICAL CDFS OF EFFORT UNDER AFFIRMATIVE ACTION



Notes: Data points are individual average efforts in one part

As can be seen, the effort of high-ability types that operated under AA policies in Part 3 is lower than those that operated under AA policies in Part 2 (both $p \leq 0.08$, FP2S test). Effort of low-ability types, however, does not differ ($p \geq 0.32$). The regression results reported in Panel (A) of Table AII.1 provide additional support for this observation: Panel (B) reports the analysis of the effect of the previous experience of the AA policy on the effort without AA.

TABLE AII.1. EFFORT AND PAST EXPERIENCE

	Panel (A): Competition under head-start and handicap				Panel (B): Competition under no affirmative action		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Experience	-2.016*	-3.337**	-3.28*	-0.695*	0.560	0.726	0.185
	(0.812)	(1.211)	(1.472)	(0.344)	(0.684)	(0.835)	(0.199)
Head-start	-0.618	-0.612	-0.56	-0.001			
	(0.823)	(0.825)	(0.974)	(0.210)			
Experience ×Head-start			-0.11		0.012	0.012	-0.096
			(0.944)		(0.799)	(0.799)	(0.208)
Low-ability	-2.818***	-4.092***	-4.09***	-0.507	-5.777***	-5.604***	-1.331***
	(0.563)	(0.710)	(0.706)	(0.283)	(0.502)	(0.473)	(0.217)
Experience × Low-ability		2.657**	2.66***	0.587		-0.333	-0.038
		(0.970)	(0.969)	(0.332)		(0.798)	(0.244)
Effort in t-1				0.815***			0.789***
				(0.034)			(0.029)
Constant	9.914***	10.453***	10.73***	2.342***	10.581***	10.483***	2.505***
	(1.950)	(1.894)	(1.903)	(0.647)	(1.628)	(1.621)	(0.567)
# Observations	1,152	1,152	1,152	1,008	1,152	1,152	1,008
# Pairs	72	72	72	72	72	72	72
Overall R^2	0.134	0.151	0.151	0.664	0.451	0.451	0.772

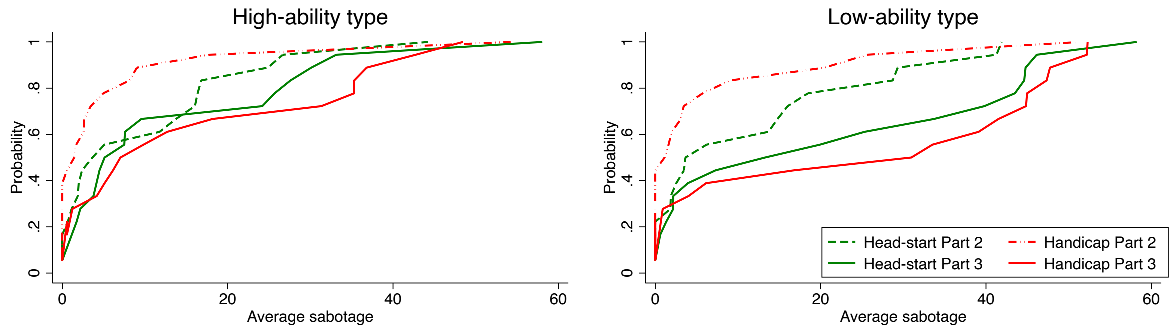
Notes: GLS panel regression (random effects) with $effort_{it}$ as the dependent variable. Robust standard errors (clustered on the pairs' level) are in parentheses. Estimations are based on the data when subjects experienced AA policies, i.e., Part 2 of the HS-NoAA and HC-NoAA treatments and Part 3 of NoAA-HS and NoAA-HC. *Experience* equals one if the data is from Part 3. *Head-start* equals one when the AA policy is head-start. *Experience ×Head-start* is their interaction. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Experience × Low-ability* is an interaction variable. *Effort in* $t-1$ corresponds to the effort in the previous period. Not reported control variables are age, rounds and a dummy for females. Controlling for lagged effort of the rival, being the tournament winner in the preceding round, and risk attitudes does not change the results. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

AII.2. Experience of Existing Affirmative Action and Sabotage Behavior

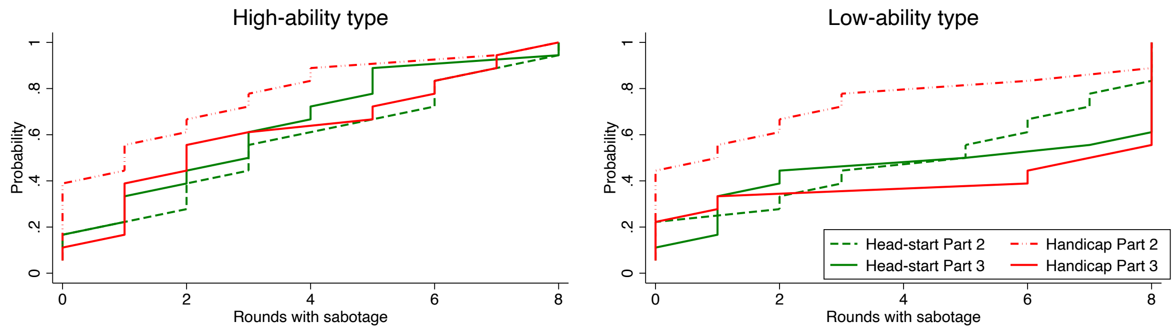
Figure 12 shows the empirical distribution of individual sabotage behavior under AA policies in Part 2 and in Part 3. It suggests that sabotage is more severe in Part 3 than in Part 2. This difference is especially pronounced for handicap (all p-values for comparison under handicap in Part 2 and Part 3 are below 0.87, with only one exception, i.e., sabotage occurrence among high-ability types, FP2S tests). Models (1)-(3) in Table AII.2 report the regression results of sabotage behavior under AA policies.

FIGURE AII.2. EMPIRICAL CDFS OF SABOTAGE UNDER AFFIRMATIVE ACTION

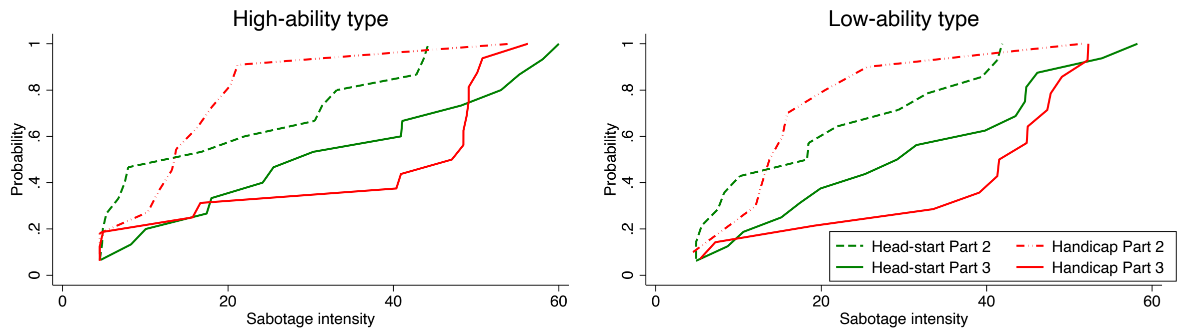
(A) Average sabotage (over all rounds)



(B) Sabotage occurence



(C) Sabotage intensity (conditional on sabotage > 0)



Notes: Based on mean individual values.

Table AII.2. Past Experience and Sabotage

	Panel (A): Competition under head-start and handicap						Panel (B): Competition under no affirmative action					
	(1)		(2)		(3)		(4)		(5)		(6)	
Independent variables	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*	Sabotage occurrence *Panel probit*	Sabotage intensity *Panel tobit*
Experience	0.133*	15.731***	0.044	15.785***	-0.039	10.113***	-0.077	-13.201***	-0.094	-11.340**	-0.080	-9.251**
	(0.072)	(3.126)	(0.070)	(4.395)	(0.075)	(3.780)	(0.307)	(4.332)	(0.361)	(5.439)	(0.104)	(4.160)
Head-start	0.108	-1.423	0.107	-1.422	0.097	-1.226						
	(0.072)	(3.141)	(0.072)	(3.142)	(0.068)	(2.660)						
Experience × HS-NoAA							0.152	13.633***	0.149	13.723***	0.138	9.848**
							(0.512)	(5.040)	(0.549)	(5.035)	(0.113)	(3.846)
Low-ability	0.201***	-0.377	0.111**	-0.319	0.119**	0.110	0.023	0.090	0.009	2.011	-0.034	-0.712
	(0.057)	(3.504)	(0.051)	(4.843)	(0.057)	(4.035)	(0.265)	(3.878)	(0.246)	(5.161)	(0.063)	(3.839)
Experience × Low-ability			0.188**	-0.110	0.258***	4.201			0.031	-3.867	0.061	-2.506
			(0.082)	(6.269)	(0.096)	(5.315)			(0.280)	(6.861)	(0.075)	(5.144)
Sabotage received in t-1					0.004*	0.574***					0.012***	0.694***
					(0.002)	(0.061)					(0.003)	(0.071)
Constant		10.858		10.846		10.355		42.587***		41.278***		38.980***
		(9.383)		(9.408)		(7.779)		(10.407)		(10.648)		(8.221)
# Observations	1,152	512	1,152	512	1,008	447	1,152	560	1,152	560	1,008	495
# Pairs	72	52	72	52	72	52	72	47	144	47	144	47
Wald Chi^2	44.169	37.457	45.422	37.456	59.741	134.789	13.785	58.350	14.707	58.704	50.761	131.026

Notes: Sabotage occurrence is analyzed with panel probit regressions with random effects. Marginal effects are reported. Sabotage intensity is analyzed with panel tobit regressions, conditioned on $sabotage_{it} > 0$ and censored below and above at 1 and 60. The first six columns are based on the data from Part 2 of the HS-NoAA and HC-NoAA treatments, and Part 3 of NoAA-HS and NoAA-HC treatments. Columns seven to twelve are based on Part 2 of the NoAA-HS and NoAA-HC treatments, and Part 3 of the HS-NoAA and HC-NoAA treatments. *Experience* equals 1 if the data is from Part 3 and zero otherwise. *Head-start* equals 1 when the AA policy is head-start. *Low-ability* is an indicator for the low-ability type (with the high-ability type as the reference group). *Experience × Low-ability* is an interaction. *Experience × HS-NoAA* is equal to one in Part 3 of the treatment HS-NoAA (when NoAA followed HS) and 0 otherwise. *Sabotage received in* $t-1$ corresponds to sabotage received in the previous round. Not reported control variables are age, round and a dummy for females. Controlling for risk attitudes and being the tournament winner in the preceding round does not change the results. Robust standard errors (clustered on pairs level) are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix III

Please download Appendix III from

https://www.dropbox.com/s/zk4r13lrxwxh19g/AA_Sabo_final_20230516_instructions_shared.docx?dl=0