Three essays in experimental economics

Economic effects of fairness perception, belief, and limited attention

Inauguraldissertation

zur Erlangung des akademischen Grades
Doctor of Philosophy (Ph.D.)
in Economics

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät
der Heinrich-Heine Universität Düsseldorf

Von:

Chi TRIEU, M.Sc.

aus Hanoi

Erste Gutachterin:
Univ.-Prof. Dr. Hannah SCHILDBERG-HÖRISCH

Zweiter Gutachter:
Univ.-Prof. Dr. Hans-Theo NORMANN

September 28, 2020

Acknowledgements

The completion of this dissertation would not have been possible without the support of many people!

First of all, I want to express my deep gratitude to my first supervisor, Hannah Schildberg-Hörisch. I was lucky to have her as my supervisor and as a co-author. Her knowledge, work ethic, and patience guided and inspired me through the most challenging time of my PhD. I sincerely thank my second supervisor, Hans-Theo Normann for fostering my interest in experimental economics and for his timely support along the way.

My gratitude extends to two of my amazing co-authors who influenced me on both an academic and personal level. I thank Jana Willrodt for meticulously reviewing my codes, slides, and writing; many times over. I have learned an immeasurable amount from her invaluable advice. I thank Ismael Martínez-Martínez for showing me the importance of properly fulfilling even the smallest task. My heartiest thanks are for his excellent advice in various matters, for the pleasant coffee breaks, and for encouraging me to believe in myself.

I am very grateful for the great conditions and the generous support at the Düsseldorf Institute for Competition Economics (DICE) which allowed me to primarily focus on research. I also thank my colleagues at DICE for plenty of fruitful discussions and debates. My heartfelt appreciation especially goes to Andrea Ciani, Fang Liu, Jannika Schad, Judith Schwartz, Maria Friese, Nesma Ali, and Vasilisa Petrishcheva for their accompany and all the fun we had over the last few years. The city of Düsseldorf definitely deserves a shout-out. It was a blessing to spend my PhD life in such a wonderful city on the beautiful river Rhine.

I owe my thanks to my parents for the countless sacrifices they have made for my sake. I thank my dear brothers for always knowing how and when to make me laugh. The acknowledgments would not be complete without mentioning Emren Düzgün. He is the most important person in my life for many reasons. This journey would not have been possible without his unlimited and unconditional support from day one!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Behavioral economics evolves around the psychological underpinnings of economic decision-making. Over the last decades, it has become an established field of economics and has shed new light on our understanding of important economic questions. Many empirical evidences that contribute to the advancement of the behavioral approach are established with data from experiments, either in the lab or in the field. Laboratory experiments and field experiments are the core elements of experimental economics, an empirical method that has been shown to be particularly suitable for testing behavioral predictions. Carefully designed experiments enable exogenous ceteris paribus manipulations over the decision-making environment. This strong asset allows rigorous measurement and causal identification of the otherwise hard to capture behavioral motives.

This dissertation, written when I was a doctoral researcher at the Düsseldorf Institute for Competition Economics, consists of three essays that lie at the intersection of behavioral economics and experimental economics. Employing laboratory experiments, my co-authors and I study three prevalent concepts in behavioral economics: fairness concerns, belief, and limited attention, each in their relevant institutional contexts. More precisely, we explore the behavioral mechanism of fairness perception and motivated beliefs under affirmative action policies, and of limited attention in the market of credence goods.

Chapter 2 examines the role of fairness perception in determining the consequences of affirmative action policies.[1] Affirmative action prevails in many important decisions such as admission to universities, hiring, and promotion even though it remains a highly controversial policy. Recent empirical studies documented both positive and negative outcomes of affirmative action, and emphasized the importance of implementation features such as providing justifications or evidence of discrimination toward the favored group. We extend

[1] Affirmative action is defined as a policy that promotes the opportunities of defined minority groups within a society to give them equal access to that of the privileged majority population (Affirmative Action, Harvard Law School Blog, accessed September 20, 2020).

the literature by providing evidence that outcomes of an affirmative action policy are shaped by how fair it is perceived. Based on theories of redistributive justice, we design an experiment that allows for manipulation over the dimension of performance that affirmative action targets. Our treatments consist of three different affirmative action policies in the form of quota rules and a baseline treatment where affirmative action is not implemented. Each quota rule favors individuals whose performance is low, either due to bad luck (discrimination), low productivity, or choice of a short working time in competitions. We document substantial heterogeneity in the fairness perception of these policies. We find that higher fairness perceptions closely correlate with a higher willingness to compete, and less retaliation against winners. No policy harms overall efficiency or post-competition teamwork. Furthermore, individuals seem to internalize the norm behind the policies that are perceived as being fairest.

Chapter 3 investigates belief as a channel through which uncertainty about the favored group influences the outcomes of affirmative action. To better capture determinants of disadvantage in an increasingly diverse population, institutions implement several affirmative action policies targeting several disadvantaged groups. This implies an uncertainty as to whether an individual who belongs to one of these groups was actually favored. In a laboratory experiment, we study how this feature affects outcomes of affirmative action in the form of quotas, and compare it with two other conditions, namely affirmative action with a certain favored group and no affirmative action. We find that when a group is favored with certainty and the group identity that triggers affirmative action is made salient, affirmed individuals are wrongly perceived as less competent, both by themselves and by others. Consequently, their willingness to compete does not increase and they are selected less for teamwork. Affirmative action with uncertain favored groups does not distort belief about competence, and thus does not induce such unintended consequences. In contrast, it increases competition entry of the affirmed groups and enhances their chances of being selected for teamwork.

Chapter 4 studies how consumers' limited attention affects outcomes in a monopolistic market of credence goods. Many crucial markets, such as healthcare, repair services, and legal services are characterized as credence goods markets. They are distinguished by a distinct feature of asymmetric information between expert sellers and customers, that is, customers cannot observe the qualities they need while sellers can. Sellers might have incentives to provide unnecessarily high or low qualities, or to charge a higher price than the provided quality. Our study is motivated by contradictions between theoretical predictions and empirical evidence on market outcomes when customers can verify the type of quality they receive, as well as recent calls for more transparency in sellers' costs in some real-world markets. While theory predicts market efficiency with equal markups for different qualities, proper provided quality and customers' maximal willingness to pay, observations from laboratory experiments yield contradicting evidence of inefficiency. Our study presents both theoretical arguments and experimental evidence that customers' limited attention to sellers' costs can be an explanation. In our experimental context, we find that when costs are made salient to customers, the market becomes more efficient. Sellers are more likely to provide sufficient quality, and prices are significantly closer to equal markups. Furthermore, we find that social preference appears to play an important role in the market outcomes.

In summary, this dissertation aims to contribute to a deeper understanding of how behavioral factors affect economic decision-making, and how they could be relevant for efficient designs of institutions.

Chapter 2

Fairness perception and consequences of affirmative action policies

Co-authored with Hannah Schildberg-Hörisch and Jana Willrodt

2.1 Introduction

In 2008, Norway was the first country in Europe to mandate a 40% quota for women in the boards of directors of publicly listed companies (Bertrand et al., 2018). Twelve years later, sixteen out of the twenty-seven member states of the European Union have implemented either voluntary or compulsory gender quotas on corporate boards (European Commission, 2019).[1] Affirmative policies targeting minorities or people with disabilities are also widespread. Already in 1950, the Indian constitution mandated affirmative action targeted at improving access to higher education, jobs, and political representation of socio-economically marginalized groups in India (Bagde et al., 2016; Pande, 2003). Nowadays in the US, many universities adopt quota policies in the admission process, targeting applicants from specific ethnic backgrounds (see, e.g., Arcidiacono and Lovenheim, 2016), while companies in Germany with more than 20 employees have to reserve 5% of positions for applicants with disabilities (SGB IX – *German Social Code, Volume 9*).

Affirmative action policies thus influence crucial decisions such as hiring, promotion, or admission to higher education by companies, governments, and universities worldwide, and, at the same time, are also highly controversial (see, e.g., Fish, 2000; Fullinwider, 2011).

On the one hand, proponents of affirmative action argue that a truly fair policy should take disadvantages (e.g., due to family background, poverty, race, and gender) into account when evaluating performance to compensate for a lack of opportunities due to discrimination, historical injustice, or the "accident of birth".[2] Hence, affirmative action policies can serve as a key tool to tackle discrimination and enhance the sense of fairness in workplace environments.

On the other hand, managers and policymakers face the arguments professed by opponents of affirmative action that decisions made under such a policy are not purely based on merit. This goes against the ideal of a "fair" policy that should select or reward the best candidates.

The underlying conflict in this debate is the collision of different fairness ideals originating from different theories of distributive justice. Each professes a distinct view on which factors of their performance individuals should be held

[1] Gender quotas are mandatory in France, Italy, Belgium, and Germany for publicly listed companies, in twelve other member states for state-owned companies, and voluntary in Greece, Slovenia, and Spain (European Commission, 2019).

[2] Another argument in favor of affirmative action policies with regard to, e.g., university admission, hiring or promotions, is the need for a more diverse student body or employee composition further up the hierarchy. This aspect is beyond the scope of this paper.

accountable for. The three stylized factors that determine an individual's performance are effort, ability, and luck (Cappelen et al., 2007, 2010).

From a *libertarian* point of view, individuals should be held accountable for all three factors, making affirmative action unnecessary. *Meritocracism*, however, posits that only personal factors, i.e., effort and ability, should be considered when assessing someone's performance, justifying affirmative action that compensates for differences in luck. According to *choice egalitarianism*, people should be held accountable only for factors within their control. This means affirmative action policies should offset bad luck and differences in ability, but not self-chosen effort. Finally, *strict egalitarianism* strives for complete distributive equality, thus calling for policies that counterbalance differences in all three factors.

In an organizational context, perception of a fair workplace seems to be one of the strongest drivers of career satisfaction and retention for employees (McKinsey & Company, 2019). Among others, fairness perceptions can affect effort provision (Charness and Kuhn, 2007; Cohn et al., 2015), physical health (Falk et al., 2017), and work morale (Deller and Sandino, 2020) of workers. Despite a growing amount of empirical evidence on the consequences of affirmative action, the perceived fairness of different affirmative action policies and the implications of those fairness perceptions for the effectiveness of affirmative action policies have barely been investigated.

In this paper, we provide evidence from a laboratory experiment that implements several affirmative action policies in the form of quota rules in a tournament setting that resembles the hiring and promotion process in real-life workplaces. We explicitly link a broad set of outcomes of those policies to their fairness perception. As a further novelty, the affirmative action policies address all three determinants of performance in a unified framework. One policy reflects meritocracism, favoring individuals disadvantaged due to persistent bad luck, an impersonal factor that resembles discrimination.[3] Another favors those disadvantaged by low innate ability (a personal, but out-of-control factor) proxied by productivity in the experimental task in which practicing does not improve performance. A third policy favors individuals whose performance is lower due to providing less effort at the extensive margin by choosing a shorter working time (an in-control factor). To our knowledge, we are the first to investigate

[3]In our design, (bad) luck is a randomly assigned characteristic and remains unchanged for each subject throughout the experiment. This characteristic resembles stable causes of discrimination such as gender, skin color, or race. Similarly, Balafoutas et al., 2016; Calsamiglia et al., 2013; Fallucchi and Quercia, 2018; Petters and Schroeder, 2020 analyze affirmative action policies that compensate for differences in a randomly assigned, exogenously given characteristic.

affirmative action policies that compensate for differences in working time or productivity – although similar policies exist outside the laboratory. [4]

In particular, we address three related research questions. Do the consequences of affirmative action policies depend on which of the three factors (bad luck, low productivity, or short working time) they compensate for? Are the three affirmative action policies perceived as differently fair? And is there a link between the perceived fairness of these policies and their consequences? Answering these questions is key to understanding the approval and the implications of different affirmative action policies.

To address these questions, we elicit individual fairness perceptions for all affirmative action schemes and consider a broad set of outcomes. First, we look at the immediate consequences of affirmative action within the tournament, namely willingness to compete and efficiency (in terms of output produced). We then explore potential spillover effects on post-competition outcomes when affirmative action is no longer in place. For that purpose, we elicit separate measures for cooperation in a team and spiteful behavior targeting those favored by affirmative action ("retaliation").

We find that none of the affirmative action policies is considered less fair than no affirmative action, documenting wide acceptance for affirmative action from a normative fairness point of view. Still, heterogeneity in the perceived fairness of the different affirmative action policies is substantial: affirmative action targeting bad luck (discriminated) individuals is perceived as fairest, followed by a policy in favor of individuals choosing a short working time, while affirmative action targeting individuals with low productivity and no affirmative action are perceived as equally and least fair.

Importantly, none of the three affirmative action policies under study harms overall willingness to compete or efficiency. However, the heterogeneity in fairness perceptions goes hand in hand with more detailed consequences of these policies. Targeted subjects are more likely to compete under a quota rule instead of working under a piece rate payment the fairer the quota rule is *generally* perceived to be. Additionally, non-targeted subjects are more willing to enter a tournament with a quota rule if they *personally* perceive the quota rule as fair.

[4]For example, the Council Directive 97/81/EC of the European Union states that part-time employees may not be treated less favorably than full-time employees (see Council of European Union, 1997). This includes their equal access to promotions (although their overall performance in terms of output is typically lower). Other policies are designed to compensate for worse performance due to low productivity. For example, students with dyslexia or physical restraints such as typist's cramps or poor eyesight can get some extra time in exams (see, e.g., Disability Rights Commission, 2007).

Regarding potential spillover effects we observe no difference between policies in post-competition teamwork. Also, no policy induces retaliation at the aggregate level. Finally, we provide first evidence on internalization of those affirmative action policies that are rated as fairest: subjects still support individuals with bad luck and short working time in post-competition interactions in which the corresponding affirmative action policies are no longer in place.

In sum, the fairness perception of an affirmative action policy seems to shape its consequences. This is an important insight for the successful implementation of such policies. For example, justifying managerial decisions to favor a specific group with evidence on existing discrimination against that group may help increase the acceptance of an affirmative action policy and, in turn, positively impact its consequences.

Our paper contributes to a growing literature researching the effectiveness of affirmative action. Schotter and Weigelt, 1992 and Calsamiglia et al., 2013 show that affirmative action in the form of bonuses or lump-sum payments for subjects who face an exogenous disadvantage in competitions can increase their performance. Many studies consider gender quotas or other preferential treatment of women in labor-market related settings and find positive overall effects (e.g., Balafoutas and Sutter, 2012; Beaman et al., 2009; Beaurain and Masclet, 2016; Niederle et al., 2013). For example, there is evidence that such policies increase women's willingness to enter competitions without discouraging men (Balafoutas and Sutter, 2012; Ibanez and Riener, 2018; Niederle et al., 2013). This is also true when gender quotas are introduced endogenously by vote (Balafoutas et al., 2016). Kölle, 2017 finds that gender quotas neither harm effort provision within teams nor the willingness to work in teams. Beside gender, several studies investigate affirmative action for members of disadvantaged castes in India (Bagde et al., 2016; Banerjee et al., 2018, 2020; Jensenius, 2015). For example, Banerjee et al., 2020 show that affirmative action boosts confidence and willingness to compete of targeted subjects, but this effect disappears when affirmative action is removed.

However, there is also evidence of adverse consequences of affirmative action (Fallucchi and Quercia, 2018; Heilman et al., 1997; Leibbrandt and List, 2018; Leibbrandt et al., 2017). In particular, in Leibbrandt et al., 2017 a gender quota turns women into the target of sabotage, thereby undermining their willingness to compete. Similarly, Fallucchi and Quercia, 2018 find that the threat of retaliation reduces competition entry of targeted subjects.

These seemingly contradictory findings bring up the question under which conditions affirmative action has adverse consequences. Answering it may provide valuable insights into how affirmative action policies that are politically desired can be implemented without causing more harm than good. The studies of Ip et al., 2020 and Petters and Schroeder, 2020 provide first related evidence. In a gift-exchange game with payoffs that depend on manager productivity, Ip et al., 2020 find that quotas for female managers decrease workers' effort when women are perceived as having lower skills than men, but not when they are discriminated against in the manager selection process. In an independent, representative survey with US citizens, approval for gender quotas for leadership positions is high when women are discriminated against in the recruitment process, but low otherwise (regardless of whether a gender skill gap exists). Petters and Schroeder, 2020 study the effect of randomly assigned quotas on peer-ratings of performance and find that targeted individuals' performance is rated worse than that of non-targeted individuals with a similar performance.

These studies indicate that the effects of affirmative action policies can depend crucially on whether and how they are justified, which in turn may impact their perception as more or less fair. This observation lays the ground for jointly studying perceived fairness and effects on outcomes of such policies. Inspired by real-world policies, criteria for affirmative action in most existing laboratory and field experiments are gender or ethnicity. However, such policies are sure to be perceived differently by different people. Take the most widely studied example of a gender quota: if a woman's performance is not among the best, some might perceive this as being the result of discrimination, while others may attribute her performance to low innate productivity, or a personal choice of working part-time. Usually, we cannot observe which of the three perceptions (or a mixture thereof) is invoked, although this is crucial to understanding the reaction to a gender quota.

To avoid this problem, we study affirmative action in a more stylized environment. In particular, we investigate the results of three different policies, each based on one of the three separate determinants of performance: persistent bad luck (resembling discrimination), effort (measured by self-chosen working time), and innate productivity. Compared to quotas for women or minorities, these affirmative action policies explicitly state the reason for a favored treatment of the respective target group. As a consequence, different judgments regarding their fairness can be unequivocally attributed to holding different fairness ideals instead of possibly reflecting different perceptions of the reasons for the target group's favored treatment. Our design thus provides a sound basis for analyzing whether and how the fairness perception of a specific affirmative action policy impacts its consequences.

Our approach thus differs from Ip et al., 2020 and Petters and Schroeder, 2020 in several respects: (i) the fact that we explicitly elicit the fairness perception of affirmative action policies, (ii) the nature of the affirmative action policies under study, and (iii) the context and outcomes we consider (willingness to compete, output produced, post-competition cooperation in teamwork and retaliation).

We introduce an experimental design allowing us to quantify productivity, working time, and luck separately and precisely within a unified framework. Each subject participates in two sessions, conducted in two consecutive weeks. We measure productivity and choice of working time in the preparatory session of the experiment. In a real effort task, we first measure subjects' productivity and their individual choice of working time. Luck is a randomly assigned multiplier, which upgrades output of lucky subjects, but downgrades output of unlucky subjects, discriminating against the latter group.

In the main session, we build on the design of Balafoutas and Sutter, 2012, yet differ significantly in the criteria for affirmative action and add several outcome variables. In our design, affirmative action policies are based on the three determinants of performance. We vary the rule determining winners of the tournament between treatments. In the control treatment, the winners are the two subjects with the highest performance. In the luck/working time/productivity treatments, at least one of the two winners must be a subject that is unlucky, has a short working time, or low productivity, respectively. We argue that differences in the consequences of affirmative action across treatments are due the criteria (luck, or productivity, or working time) this policy is based on.

Our study is the first to analyze the consequences of affirmative action policies related to productivity and working time. We are also able to systematically compare affirmative action policies based on the three determinants of performance in a unified framework. Moreover, we provide novel evidence on heterogeneity in the perceived fairness of affirmative action policies and link fairness perceptions to its consequences. Our finding that the fairness perception of an affirmative action policy can shape its consequences is key for the communication and successful implementation of such policies if they are politically desired.

The remainder of the paper is organized as follows. Section 2.2 explains our experimental design, while section 2.3 presents our results on the consequences of affirmative action on willingness to compete, efficiency, cooperation, and retaliation and provides evidence that subjects internalize affirmation action policies beyond the context in which they are applied. Section 2.4 discusses our findings and concludes.

2.2 Experimental design

Our experiment employs a combination of a within- and between-subject design, in which the four treatments are assigned across subjects: one control treatment without affirmative action and three treatments with different affirmative action policies. Each affirmative action policy favors subjects with a characteristic that dampens their performance – either subjects who have bad luck (which resembles discrimination), those with low productivity, or those who have chosen a short working time.

Each subject participates in two sessions taking place in consecutive weeks: a preparatory session and the main session. The purpose of the preparatory session is to learn about each subject's productivity and individually chosen working time to classify them into high and low productivity, high and low working time types, respectively. This determines which subjects will be favored by the respective affirmative action policies. In the main session, we investigate the fairness perception of the various affirmative action policies and their consequences on willingness to compete, efficiency, and post-competition cooperation and retaliation. Table 2.1 provides an overview of the experimental design.

TABLE 2.1: Summary of experimental design.

Preparatory session		Practice round (grid task)
		Measurement of baseline productivity (grid task)
		Questionnaire
		Measurement of choice of working time (grid task)
Main session	Stage 1	Piece rate (grid task)
	Stage 2	Tournament (grid task)
	Stage 3	Choice between piece rate and tournament (grid task)
	Stage 4	Group work (slider task)
	Stage 5	Dictator Game
		Measurement of fairness perception

2.2.1 The real-effort task

We apply the different affirmative action policies to performance in a tournament that is based on what we call the *grid task*, a real-effort task introduced by Abeler et al., 2011. Subjects work on this task several times under different incentive schemes (see Table 2.1). In this task, subjects count the number of zeros in a 10-by-10 table containing 100 digits of randomly distributed zeros and ones (see Figure 2.1).

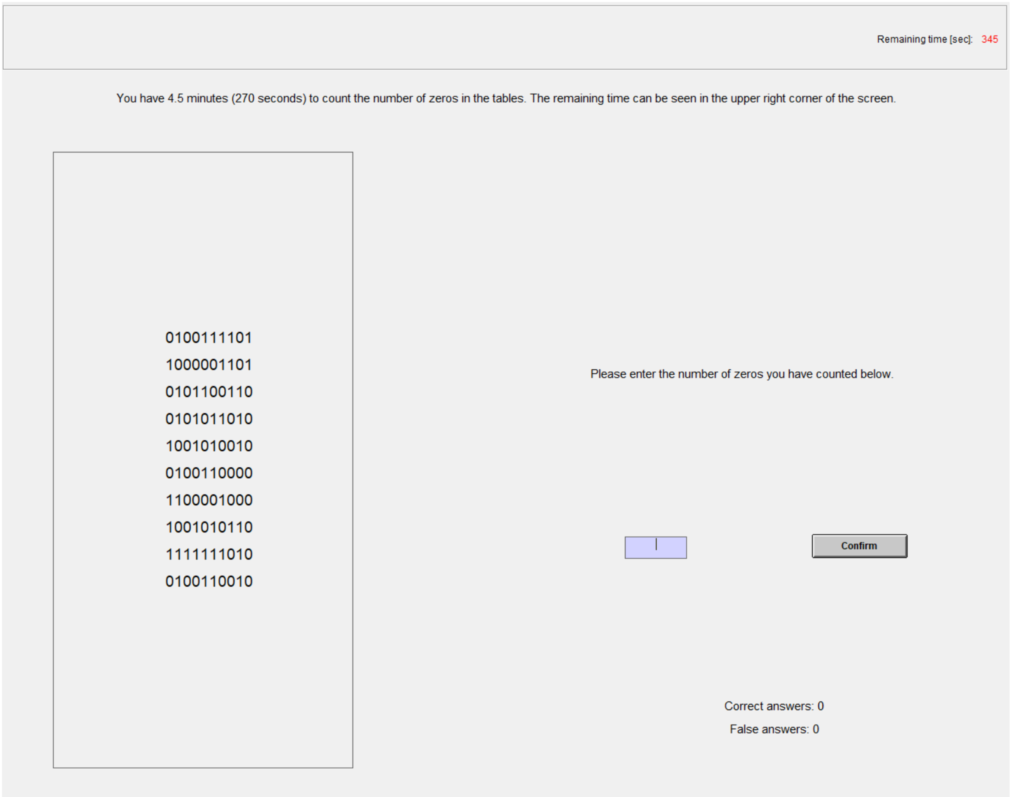

FIGURE 2.1: Exemplary screen of grid task

The grid task has several desirable attributes. First, the tediousness of the task induces a positive effort cost and minimizes experimenter demand effects (Abeler et al., 2011). Therefore, we are confident that our measure of working time (the time subjects decide to work on this task) actually captures the effort subjects are willing to spend on the task. Second, our data show substantial variation in productivity and chosen working time for this task (see Figure 2.7 and Figure 2.8 in section 2.A). Third, the grid task does not require special prior knowledge or skills. Moreover, as Balafoutas et al., 2016 note and our data confirm, the grid task is a gender-neutral one. For example, the average number of correctly solved grids in the five minute grid task in the preparatory session is 7.46 for men and 7.69 for women (Mann-Whitney U test, $p = 0.232$).[5] Most importantly, the task allows us to clearly distinguish between the three determinants of performance that matter for perceived distributional fairness according to different fairness ideals (see, e.g., Cappelen et al., 2007).

2.2.2 Preparatory session

The preparatory session consists of four parts (see Table 2.1). First, subjects familiarize themselves with the grid task in a practice round of two minutes. We then measure each subject's productivity, followed by a questionnaire and a choice of individual working time. The purpose of this is to classify subjects into binary types with high and low productivity or long and short working time, respectively. Depending on the treatment allocation, a subject's type determines whether a subject will be among those favored by the affirmative action policy in the main session.

Classification into productivity type: Subjects are asked to solve as many grids as they can within five minutes at a piece-rate of 0.50 EUR for each correctly solved grid. Subjects who solve more grids correctly than the median are classified as being of the high productivity type, while those below the median are classified being of the low productivity type.[6] Figure 2.7 in section 2.A in the appendix displays the distribution of the number of correctly solved grids in this stage.

Classification into working time type: At the end of the preparatory session, subjects work on the grid task for another time, now at a piece-rate of 0.10 EUR per correctly solved grid. Subjects can now freely choose how long they want to work. After every grid, subjects can choose to continue or stop working by clicking on the corresponding button. If subjects choose to stop working, they finish the preparatory session and can leave the laboratory immediately. To minimize peer effects in the decision when to stop working, we implement a flexible show-up policy, meaning that subjects start the session individually and do not

[5] Considering only those subjects who participated in both sessions, as we do in section 2.3, these numbers hardly change (7.54 and 7.68 correctly solved grids for men and women, respectively; not significantly different according to Mann-Whitney U test: $p = 0.478$).

[6] Our measurement of productivity (number of correctly solved grids per minute) might capture both the given ability of working on the grid task and effort at the intensive margin. However, subjects tend to exert maximum effort in laboratory real effort tasks with short working period (Araujo et al., 2016; Corgnet et al., 2015; Gächter et al., 2016; Goerg et al., 2019). Therefore, our preferred interpretation of productivity is that it reflects ability at the grid task, which is beyond subjects' control during the experiment.

reach this last stage at the same time.[7] We truthfully communicate to subjects upfront that their chosen working time has additional consequences on the session in the following week. Based on whether their chosen working time is below or above the median, they are classified as being either of the short or long working time type. This determines how long they will work on the grid task in the main session and has consequences on their expected earnings in the main session.[8] The low piece-rate was deliberately chosen to make the task less attractive so subjects would choose to stop working after a reasonable time. On average they do so after 24.13 minutes.[9] Figure 2.8 in section 2.A displays the distribution of working time. In our data, the number of correctly solved grids in the first stage and chosen working time in the last stage are not significantly correlated (Pearson's correlation coefficient: $\rho = 0.061$, $p = 0.192$).[10]

Questionnaire: We elicit a number of control variables, including measures of risk and social preferences, cognitive ability (Raven matrices), personality (Big Five), and socio-demographics. Section 2.D provides more details on the questionnaire.

2.2.3 Main session

The main session consists of five stages and a final questionnaire.[11] In Stages 1 to 3, subjects work on the grid task repeatedly, with their payoff-relevant performance being determined as follows:

[7] In the recruitment e-mail, subjects are informed that they can show up at the lab at any time within a two-hour interval.

[8] Refer to subsection 2.C.1 in the appendix for details on how the consequences on the working time decision were communicated. In particular, the instructions state "Your working time today will determine your working time in the next session next week. Next week, you will work on a similar task and you will be given a specific amount of time to solve as many tables as possible and get paid accordingly. (...) Based on your chosen working time today, we will form two groups. One group contains that half of the participants who choose to work for a shorter time today. This group will also be given a shorter time to work in the session next week. The other group contains that half of the participants who choose to work for a longer time today. This group will also be given a longer time to work in the session next week. Who works shorter will, on average, solve less tables correctly and therefore earn less. The experiment is, however, shorter (it will end earlier). Who works longer will, on average, solve more tables correctly and therefore earn more. The experiment is, however, longer (it will end later)."

[9] Considering only those who participated in both sessions the average time after which subjects stop working is 24.17 minutes.

[10] Considering only those who participated in both sessions we find a correlation coefficient of $\rho = 0.034$ ($p = 0.513$).

[11] Part of the design of the main session builds on Balafoutas et al., 2016. We thank the authors for sharing their ztree program and instructions with us.

$$\textbf{Performance = Correct grids per minute} \times \textbf{Working time} \times \textbf{Luck multiplier}$$

Assignment of types. In our experiment, each subject is of one of eight (2^3) types: high or low productivity × long or short working time × lucky or unlucky. Subjects are fully informed about all three dimensions of their own type before they enter the first stage of the main session and a subject's type stays constant throughout the experiment. The *productivity* and *working time type* are assigned to each subject based on the outcomes of the preparatory session as described in subsection 2.2.2 above. While productivity is something given, subjects are given more or less time to work on the grid task in the main session according to their chosen working time in the preparatory session. Subjects of the long working time type are given 7.5 mins to work on the task, while those of the short working time type have only 4.5 mins. *Luck* is reflected by a randomly assigned multiplier. Half of the subjects are *lucky*. They are assigned a high multiplier of 1.25. The other half are *unlucky* being assigned a low multiplier of 0.75. The number of correctly solved grids in the total time worked is weighted with multiplier.[12] Those parameters were chosen to make the effects of each policy on the probability of winning for those favored (not favored) by it comparable in size.[13]

Stage 1: Piece Rate. This stage provides a baseline measure of performance without tournament incentives. Subjects work on the grid task according to their type's working time and receive a piece-rate payment of 0.50 EUR for each correctly solved grid multiplied with the respective luck multiplier.

Stage 2: Tournament. In stage 2, subjects solve the grid task under tournament incentives, each competing against five other subjects.[14] The purpose of

[12]For example, if a subject has solved 12 grids correctly and has been assigned the high (low) multiplier, the 12 grids are treated as 15 (9) grids. By randomly assigning the luck multipliers at the beginning of the main session, and having them remain constant for each subject throughout the experiment, we effectively discriminate against unlucky individuals.

[13]Specifically, after having observed the productivity distribution in the preparatory session, we run a simple simulation that determines how likely it is on average for the low and high productivity type to win the tournament with and without affirmative action regarding productivity. We then choose parameters for the working time of the short and long working time type such that the change in the average probability of winning the tournament when introducing affirmative action regarding working time is comparable. This implies that the luck multiplier for unlucky and lucky subjects is chosen such that the change in the average probability of winning the tournament when introducing affirmative action regarding luck is also comparable.

[14]A potential confound of the tournament outcome could arise from an unbalanced composition of types across groups because the probability of winning the tournament depends both on one's own and others' performance, which in turn is affected by the types of competitors. To eliminate this effect, each group of six consists of three subjects with low productivity, three with high productivity, three with long working time, three with short working time, three with the high and three with the low luck multiplier. At the beginning of stage 2, subjects are informed

this stage is to measure the effect of different affirmative action schemes on performance. Among each group of six, the two winners of the tournament receive 1.50 EUR per correctly solved grid multiplied with their luck multiplier, while losers receive nothing, keeping the average payment constant compared to stage 1. If necessary, a random tie-breaking rule is applied to determine winners. Winners and rank within each group of six are not announced until the end of stage 5. How exactly the winners are determined is the main treatment variation of our experiment.

Between-subject treatments. We conduct four treatments: one control treatment without affirmative action and three different affirmative action (AA) treatments. In the control treatment without any quota rule, the two subjects with the highest performance are the winners. In the affirmative action treatments, a quota rule is added to determine the winners. If this rule is not automatically fulfilled, the subject with the second-highest performance is replaced by the highest-performing subject who fulfills the quota criterion. The following quota rules apply in the treatments:

- *Control (CTR)*: No quota rule.
- *Affirmative action w.r.t. luck (AAL)*: At least one subject of the unlucky type has to be among the two winners.
- *Affirmative action w.r.t. working time (AAW)*: At least one subject of the short working time type has to be among the two winners.
- *Affirmative action w.r.t. productivity (AAP)*: At least one subject of the low productivity type has to be among the two winners.

Control questions make sure that subjects understand the tournament scheme before starting to work on the task.[15]

Stage 3: Self-selection into tournament. To elicit the willingness to enter the tournament, subjects work on the grid task again. In this stage, they choose whether they would like to work under piece-rate incentives (exactly as in stage 1) or tournament incentives (as in stage 2). Importantly, if a subject chooses the tournament in *stage 3*, her performance will be compared to the performance of

about this rule for group composition, but not about the specific type of other each group member. Group composition remains the same in all following stages.

[15] After having read the rules of stages 2 and 3, subjects have to answer control questions correctly before they can start working on the grid task. These multiple-choice questions describe scenarios about competition within a group, provide information about each member's performance and who is favored by affirmative action, and ask about the winners. The control questions cover both cases in which affirmative action does or does not change the results of the competition.

her five fellow group members in *stage 2*. This feature ensures that a subject's decision to enter the tournament is independent of her belief about others entering the tournament (compare Niederle et al., 2013).

Belief elicitation. At the end of stage 3, subjects report their beliefs about their relative performance in stage 1, 2, and 3. Subjects are asked to guess their rank both within the whole group of six and within the group of three subjects with the same luck type (in treatment AAL), the same working time type (in treatment AAW), or the same productivity type (in treatment AAP). One guess is randomly chosen to be payoff-relevant. Subjects receive 1 EUR if they guess correctly.

Stage 4: Cooperation in group work. Stage 4 keeps the group composition and treatment history from previous stages, but provides a new working environment with a new task and new payoff rules. Compared to previous stages, all subjects now work for the same amount of time (5 mins), and there are neither multipliers nor affirmative action.

In the slider task (Gill and Prowse, 2012), subjects are shown a series of screens, each with 6 sliders on them. Each slider has a range of positions between 0 and 100. Sliders are solved by using the computer mouse to move the slider markers to the position of 50.[16] A screen is considered "solved" if all six sliders are positioned at 50. Only then can a subject continue to the next screen.

Importantly, each correctly solved screen yields 0.60 EUR *for the group as a whole*, 0.10 EUR for each of its members. Since all group members benefit from an individual's effort, this is a typical setup to measure cooperation and how much a subject works indicates their willingness to contribute for the benefit of the group. In order to keep the previous tournament experience with or without affirmative action salient, we introduce unequal bonuses for winners and losers of the tournament in stage 2 (as Balafoutas et al., 2016, do). Subjects receive a bonus of 5 EUR if they were the winners in stage 2, and 2 EUR otherwise.

Stage 5: Dictator Game. Subjects play one Dictator Game with each of their five group members. The only thing they know about the other group members is whether they were winners in stage 2 and whether they were favored by affirmative action. All five Dictator Games are displayed on the same screen. For each game, subjects are endowed with 5 EUR, and can decide how much to give away in 0.1 EUR increments. This setup is used to learn how more or less favorably subjects treat specific other subjects after the tournament phase.

[16]To make sure that subjects use only the computer mouse to solve the task, the left and right arrow keys of the keyboard are disabled.

Fairness perception. After stage 5, subjects are asked how fair they perceive the different policies to be. They first rate the policy in their own treatment on a seven-point Likert scale, then the policies that appear in the other treatments. Thus, we assess the fairness perception of each subject for each of the four policies. The fairness questions describe the policies neutrally and do not mention the term "affirmative action" (see section 2.B for the exact wording).

2.2.4 Procedures

The experiment was conducted at the DICE Lab at the University of Düsseldorf in April 2018 and the BonnEconLab at the University of Bonn in August 2018 using the software zTree (Fischbacher, 2007). On average, each session lasted 90 minutes. Subjects were recruited via ORSEE (Greiner, 2004) and Hroot (Bock et al., 2014) from the subject pools of the respective labs, both of which include students of various disciplines. In the recruitment email, subjects are informed that the experiment consists of two mandatory sessions and that all payments will only be realized at the end of the second session. Only 7 out of 463 subjects who participated in the first preparatory session did not show up in the second one (main session), implying an attrition rate of 1.5%. To be able to match the data of both sessions while ensuring anonymity, we asked subjects to generate an ID (that is never connected to their name) at the beginning of the preparatory session, and to re-enter it in the main session. In total, the number of subjects in each treatment is 108, 84, 90, and 90 for CTR, AAL, AAW, and AAP, respectively.[17] Throughout the paper, we focus our analysis on those subjects who participated in both sessions unless explicitly stated otherwise. Gender composition does not vary significantly between treatments (59% females; Kruskal-Wallis test: $p = 0.582$).

On average, subjects earned 26 EUR for both sessions. The payoff of the preparatory session is the sum of payoffs from the measurements of productivity and choice of working time in the grid task, a risk choice list, and a fixed payment of 2 EUR for completing the questionnaire. The payoff of the main session consists of a 4 EUR show-up fee and the earnings from one randomly chosen stage. Subjects know that one of the five parts of the main session will be randomly chosen to be paid. Instructions are distributed stage-by-stage. At the end of each of the first four stages, subjects are informed about their individual performance. Subjects never learn the performance of other subjects.

[17]To guarantee a similar composition of all groups (see description of Stage 2: Tournament), we have to exclude around 18% of subjects who showed up from participating in the main session.

2.3 Results

We start by presenting evidence on the fairness perception of the various affirmative action policies. We then analyze the consequences of affirmative action for willingness to compete in the tournament and efficiency, before we examine its impact on post-competition cooperation and retaliation. Finally, we provide evidence that affirmative action policies are internalized beyond the context in which they are binding. Throughout, we discuss links between fairness perception and consequences of affirmative action policies.

2.3.1 Fairness perceptions of affirmative action

Based on fairness considerations, there is rather broad support for all affirmative action policies under study (see Figure 2.2 and Figure 2.9 in the appendix for the distributions of answers). In particular, no affirmative action policy is considered less fair than the absence of affirmative action in the control treatment. While the control treatment is perceived as somewhat unfair (only 19% assign a rating at or above 5 on a Likert scale from 1 to 7; average rating 3.25), affirmative action in favor of unlucky subjects (AAL) is perceived as fairest (average rating of 4.70, with 48% rating it at or above 5). This is followed by affirmative action in favor of subjects of the short working time type (AAW; average rating 4.23, 34% rating it at or above 5). Affirmative action favoring subjects of the low productivity type (AAP) is on average rated roughly equal to no affirmative action (average score for AAP 3.31) but leads to more heterogeneity in fairness rating than CTR (33% rate AAP at or above 5). Wilcoxon signed rank tests for all pairwise policy comparisons yield $p < 0.001$.[18] The only exception is the absence of a significant difference in perceived fairness of treatments CTR and AAP (Wilcoxon signed rank test, $p = 0.327$). Interestingly, the ranking of treatments by perceived fairness diverges slightly from meritocracism, according to which a policy controlling for an out-of-personal-control factor (productivity) would be rated as fairer than an in-control factor self-chosen working time.

Result 1 (Fairness perception): On average, affirmative action policies that compensate for bad luck (discrimination) or short working time are perceived as significantly fairer than no affirmative action. Affirmative action based on low productivity is perceived as equally fair as no affirmative action.

This pattern of fairness perceptions is similar across treatments (see Figure 2.10 in the appendix). Moreover, the average fairness rating of favored and

[18]Throughout the paper, we report two-sided tests unless explicitly stated otherwise.

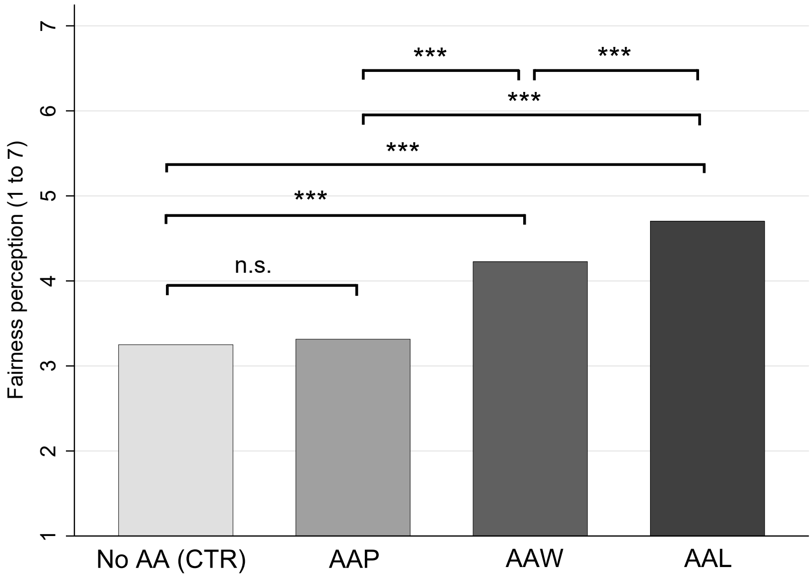

FIGURE 2.2: **Fairness perception**

Notes: Fairness perceptions for different policies. Higher numbers indicate that a policy is perceived as fairer. The brackets and stars above each bar show results of Wilcoxon signed rank tests. *** indicates $p < 0.001$.

unfavored subjects does not differ significantly.[19] Importantly, these findings jointly underline that fairness judgments regarding affirmative action are not strongly shaped by individual experiences.

2.3.2 Willingness to compete

A key purpose of affirmative action policies is to encourage those favored by them to enter the competition. In this section, we will analyze whether the different policies indeed affect subjects' willingness to select into the tournament when given the choice between working under a piece rate and a tournament scheme.

In section 2.3.2 we start by comparing the average willingness to compete in stage 3 across treatments and then take a closer look at favored and non-favored types. Throughout, we will point out close links between fairness perceptions and willingness to compete. In section 2.3.2 we consider determinants of willingness to compete at the individual level.

[19]An OLS regression of all 1056 fairness ratings from 264 subjects in the three treatments with affirmative action on a binary variable indicating whether a subject is favored yields a coefficient of -0.072 with a p-value of 0.603 (standard errors clustered at subject level).

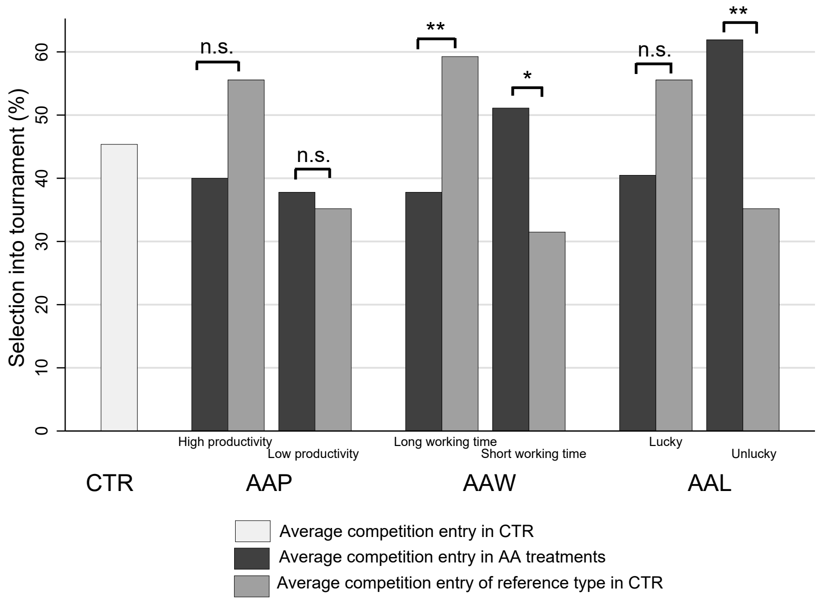

FIGURE 2.3: **Willingness to compete by treatment and type**

Notes: Proportion of competition entry in stage 3 of a given type in the treatments with and without affirmative action concerning their type. The brackets and stars above each bar show results of Fisher's exact tests, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The big picture: willingness to compete at the aggregate level

Overall, the proportion of subjects selecting the tournament differs modestly across treatments. 45% of subjects are willing to compete in the absence of affirmative action (CTR) compared to 39% in AAP, 44%, in AAW, and 51% in AAL. For the affirmative action schemes, the overall pattern mimics that in Figure 2.2: subjects tend to be more willing to compete under affirmative action schemes that are generally perceived as fairer. However, differences in willingness to compete are not significant across treatments.[20]

Since ultimately, we want to learn about the effects of affirmative action on those who are favored by it compared to those who are not, this result warrants closer inspection. We thus continue by comparing the willingness to compete of each type (e.g., the unlucky type) in the treatment that concerns them (e.g., AAL) to that of the same type in the control treatment (see Figure 2.3).

We observe a similar tendency in AAL and AAW. While affirmative action increases favored subjects' willingness to compete, it tends to lower tournament entry of non-favored subjects. The encouragement effect is most striking for AAL, in which unlucky subjects' tournament participation increases by 26.7 percentage points compared to CTR (Fisher's exact test, $p = 0.013$). Subjects of the

[20]Fisher's exact tests for pairwise comparisons with the control treatment yield $p = 0.388$ for AAP, $p = 1.000$ for AAW, and $p = 0.468$ for AAL. Likewise, a Kruskal-Wallis test does not detect any significant differences across all four treatments ($p = 0.446$).

short working time type are 19.6 percentage points more likely to compete in AAW than in CTR (Fisher's exact test, $p = 0.064$). However, in both treatments, we also observe a discouragement effect of 15.1 and 21.5 percentage points, respectively, on non-favored subjects (not significant in the case of AAL; Fisher's exact tests, $p = 0.156$ for AAL and $p = 0.044$ for AAW).

In contrast, in AAP tournament entry of low productivity subjects hardly increases compared to CTR (2.6 percentage points; Fisher's exact test, $p = 0.836$), while high ability subjects tend to be discouraged from entering (15.6 percentage points, Fisher's exact tests, $p = 0.158$).

Interestingly, the encouragement effect of affirmative action on the favored subjects closely mirrors the average fairness rating of the affirmative action treatments (compare Figure 2.2) with higher encouragement in treatments that are generally perceived as fairer.

Result 2a (Willingness to compete): The higher the average fairness rating of an affirmative action scheme, the more willing are favored subjects to compete.

Result 2a establishes a relationship between the average fairness perception and the effectiveness of affirmative action to increase its target group's willingness to compete. A higher average fairness perception of an affirmative action scheme goes hand in hand with a higher willingness to compete of the favored subjects. A plausible explanation for such a relationship is that subjects anticipate the social acceptance, i.e., the average fairness perception of a given quota rule, and feel more comfortable to enter a competition, in which they are favored if it is generally judged as fair. Of course, average perceived fairness of an affirmative action scheme also reflects personal perceived fairness of that scheme. In the following, we will investigate the relationship between fairness perception and willingness to compete at the individual level.

Determinants of the willingness to compete at the individual level

Table 2.2 reports the marginal effects resulting from probit regressions on the determinants of willingness to compete at the individual level. The dependent variable is an indicator variable that equals 1 if a subject chooses to enter the tournament and 0 if they choose the piece rate payment scheme in stage 3. From a theoretical perspective, an individual's belief on the likelihood of winning the tournament and risk attitude should be the key determinants of this choice. Empirically, there is also substantial evidence that women are less likely to compete (Almås et al., 2016; Buser et al., 2014; Croson and Gneezy, 2009; Gneezy

TABLE 2.2: Individual level determinants of competition entry

	(1)	(2)	(3)
Belief of rank	-0.100***	-0.088***	-0.119***
	(0.015)	(0.017)	(0.017)
Risk attitude	0.026**	0.029**	0.031***
	(0.010)	(0.012)	(0.012)
Female	-0.093*	-0.104*	-0.091*
	(0.049)	(0.057)	(0.055)
Fairness perception	0.019	0.028*	0.049**
	(0.013)	(0.015)	(0.021)
Favored			0.480***
			(0.129)
Fairness perception × Favored			-0.067**
			(0.029)
N	372	264	264

Notes: Average marginal effects from a probit regression. Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The binary dependent variable is willingness to compete in stage 3 (1 if tournament, 0 if piece-rate). *Belief on rank* is a categorical variable about beliefs regarding own rank (between 1 and 6) in stage 3; *Risk attitude* is the answer to the general risk question elicited on an 11-point scale, higher numbers indicate a higher willingness to take risks; *Female* is an indicator variable for gender (1 if female, 0 if male); *Fairness perception* reflects fairness rating of own treatment, elicited on a 7-point scale on which higher numbers indicate higher perceived fairness; *Favored* is an indicator variable (1 if favored, 0 otherwise).

et al., 2003; Leonard et al., 2009; Niederle and Vesterlund, 2007, 2011; Sutter and Glätzle-Rützler, 2015). Given the results in section 2.3.2 we are interested in the role of the fairness perception of the policy in place.

Results in column (1) imply that the belief regarding their own rank in the tournament, risk attitudes, and gender are indeed predictive for individual willingness to compete. If a subject's belief about their own rank increases by 1 (on a scale from 1 to 6, with 1 as the highest rank), it is about 10 percentage points less likely to enter the tournament.[21] If their willingness to take risks increases by 1 (on an 11 point Likert scale), they are 3 percentage points more likely to

[21] Table 2.4 replicates all specifications of Table 2.2 and shows the marginal effect of each value of belief separately. Moreover, results are qualitatively the same and quantitatively very similar if we use beliefs regarding own rank in stage 2 instead of stage 3.

compete. Women tend to be 9 percentage points less likely to compete compared to men. Finally, on average an individual's own fairness perception is not a significant predictor of willingness to compete. We move on by investigating whether the impact of fairness perceptions differs for favored and non-favored subjects (see column (3)). Since this requires dropping subjects from the control treatment, column (2) serves as a reference, replicating column (1) with subjects in the affirmative action treatments.

Results in column (3) imply that an individual's own fairness perception is a highly significant and influential predictor of the non-favored subjects' willingness to compete: if their fairness perception of a given affirmative action treatment increases by 1 (on a 7 point Likert scale), they are about 5 percentage points more likely to compete. Moreover, the specification reported in column (3) reveals that favored subjects tend to have a higher willingness to compete but their individual fairness perception of an affirmative action scheme does not have any effect on their willingness to compete. A test of joint significance of Fairness perception and Fairness perception $\times$ Favored yields $p = 0.429$. Finally, women are slightly less likely to compete than men.

Result 2b (Willingness to compete): At the individual level, willingness to compete is driven by an individual's belief of winning and risk attitude. Moreover, non-favored subjects are more likely to compete, the fairer they consider an affirmative action scheme to be, while this is not the case for favored subjects.

Overall, a higher fairness perception of affirmative action schemes increases the share of competing subjects in two ways: a higher *general* fairness perception of a given quota rule goes hand in hand with a higher willingness to compete of favored subjects and a higher *individual* fairness perception raises non-favored subjects' willingness to compete.

While encouraging favored individuals to enter competitions is the key aim of affirmative action, raising non-favored subjects' willingness to compete is also desirable since competition tends to increase efficiency (see, e.g., Balafoutas et al., 2016, Niederle and Vesterlund, 2007, and Figure 2.4). Thus, our results point at an important aspect in designing and communicating affirmative action schemes that has so far been disregarded, namely their fairness perception. For an affirmative action policy to encourage favored subjects to compete without discouraging non-favored subjects, it is vital that the policy is perceived as fair in general (by others) and personally. So from a policy perspective, providing a convincing rationale for the implementation of quota rules to ensure that they are largely perceived as fair seems key to make them a success. An example would be providing evidence on the discrimination against women when introducing a gender quota (Ip et al., 2020).

2.3.3 Efficiency

A prominent worry of opponents of affirmative action is that it harms efficiency, i.e., that it results in not selecting or rewarding the "best". A 50% quota rule that replaces the second-highest performer as a winner by someone else may seem damaging to efficiency from an ex-ante perspective. But at closer inspection this is not necessarily true.

To learn about efficiency we use the tournament winners' number of correctly solved grids per minute as a measure, thus not considering differences in performance that arise due to luck or working time. This measure actually captures two separate aspects. The first is who the winners are and would be of most interest in a promotion or selection context. This is determined both by winner selection according to the quota rule (or its absence) and participants' entry decision. The second determinant is how well winners perform under the given incentives. This is likely driven by motivation and more important when output during the tournament matters.[22]

Qualitatively, results are the same, no matter whether we consider the number of correctly solved grids per minute of stage 3 winners in stage 3 or in the preparatory session as an efficiency measure (see Figure 2.4): the number of correctly solved grids per minute does not differ significantly between the control treatment and those with affirmative action (test results reported in the figure notes).

Result 3 (Efficiency): None of the affirmative action policies under study harms efficiency.[23]

Extending the efficiency analysis to the performance of both winners and non-winners, we also observe that none of the affirmative action policies harms efficiency. These results extend the findings of Balafoutas and Sutter, 2012 and Niederle et al., 2013 that a 50% quota rule favoring women does not harm efficiency to 50% quota rules that favors individuals with bad luck, short working time or low productivity, i.e. it holds regardless of which determinant of lower

[22]For a more detailed discussion of these separate aspects refer to subsection 2.A.4 in the Appendix.

[23]Extending the efficiency analysis to the performance of both winners and non-winners, we also observe that none of the affirmative action policies harms efficiency.

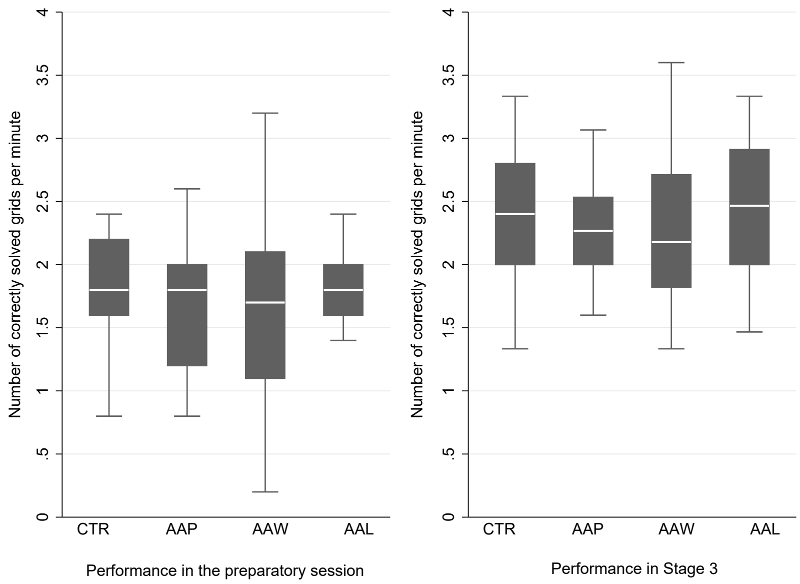

FIGURE 2.4: **Number of correctly solved grids per minute in the preparatory session and stage 3 of winners in stage 3**

Notes: Number of correctly solved grids per minute of stage 3 winners in the productivity task of the preparatory session and stage 3 for each treatment. Two outliers from CTR and AAP in the right boxplot are excluded. The upper (lower) hinges of the boxes show the 75th (25th) percentiles, the white lines inside the boxes show the median values, and the upper (lower) adjacent lines show the maximum (minimum). Stage 3 winners solved on average 1.83 grids per minute correctly In the preparatory session in the control treatment compared to 1.85, 1.73 and 1.76 in treatments AAP, AAW, and AAL, respectively. The corresponding p-values of Mann-Whitney U tests comparing the affirmative action treatments to the control treatment are 0.815, 0.554, and 0.953, respectively. In stage 3, the winners solved 2.47 grids per minute correctly in the control treatment compared to 2.51, 2.32, and 2.41 in treatments AAP, AAW, and AAL, respectively (p-values of Mann-Whitney U tests are 0.642, 0.390, and 0.920, respectively).

performance affirmative action compensates for. The fact that affirmative action does not harm efficiency even when one of the winners is replaced by a low productivity subject may be explained by the positive influence of affirmative action on targeted subjects' motivation. For example, low productivity subjects tend to solve on average 1.97 grids correctly per minute in AAP compared to 1.70 in other treatments (stage 2 performance; Mann-Whitney U test $p = 0.135$).

Finally, how fair an individual perceives the assigned treatment to be does not correlate with the number of correctly solved grids per minute in stage 3, neither for winners of the tournament nor for subjects as a whole ($\rho = -0.073$, $p = 0.554$ for winners, $\rho = -0.037$, $p = 0.482$ for all subjects).

2.3.4 Post-competition cooperation, retaliation, and internalization of social norms

Cooperation. We measure cooperative behavior by performance in the slider task (number of correctly solved slider screens) in stage 4 in which a higher individual performance yields equal benefits for all members of a group. Overall, average post-competition cooperation at the group level does not differ significantly across treatments (Kruskal-Wallis test, $p = 0.308$). Neither do we find significant differences for pairwise comparisons between cooperation levels in each affirmative action treatment and the control treatment (see Table 2.5 in the appendix; Mann-Whitney U tests $p > 0.10$ for all). Importantly, the lack of significant differences is not due to a lack of statistical power. Our sample size is able to detect an effect size as small as 0.87 (roughly a 12% performance change based on the control treatment) at the conventional level of power of 80% in each treatment using a t-test and a significance level of 0.05. This confirms the results of Balafoutas et al., 2016, Sutter et al., 2016 and Kölle, 2017 who find no spillover effects of quotas on subsequent teamwork in affirmative action contexts that differ from the context we study.

Furthermore, we find no evidence that fairness perceptions affect cooperation post-competition (Pearson's correlation coefficient between fairness perception of own treatment and performance in the slider task: $\rho = 0.047$, $p = 0.382$).

Retaliation. In contrast to the slider task, in which an individual's behavior affects all other group members equally, decisions in stage 5's dictator games allow subjects to treat each individual other group member in a more or less favorable way. They may condition their transfer on whether someone was a winner or loser and favored or non-favored by the respective affirmative action policy in the tournament in stage 2.[24]

In particular, non-favored subjects may have an urge to treat formerly favored subjects in a less advantageous manner than non-favored ones ("retaliation"). Our analysis focuses on the most pointed situation, in which such retaliation seems most likely to occur: we investigate whether non-favored losers give less to favored winners than to non-favored winners in the affirmative action treatments (see Figure 2.5). We find no evidence for such retaliation under any of the three affirmative action policies (Wilcoxon signed rank tests yield $p = 0.198$ for AAL, $p = 0.906$ for AAW , and $p = 0.317$ for AAP, respectively). Figure 2.11 in the appendix provides additional evidence on the absence of retaliation based on a broader set of situations.

[24]Considering the overall average transfer in the dictator games as an indicator, we do not observe any significant differences in pro-social behavior across treatments.

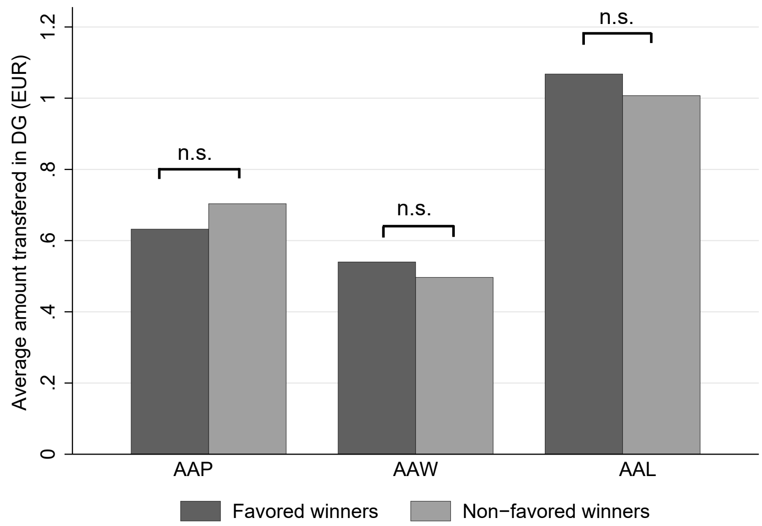

FIGURE 2.5: **Retaliation in dictator games**

Notes: Average transfer (EUR) from non-favored losers to favored winners versus non-favored winners in dictator games for each treatment. The brackets and stars above each bar show results of Wilcoxon signed rank tests, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.3: Transfer in dictator games and fairness perception

from	Panel 1: Non-favored losers				Panel 2: All subjects			
to	fw	nfw	fl	nfl	fw	nfw	fl	nfl
Fairness	0.262**	0.211*	0.101	0.089	0.203**	0.099	0.044	0.018
	(0.122)	(0.120)	(0.102)	(0.090)	(0.083)	(0.079)	(0.063)	(0.058)
Constant	-0.966*	-0.736	0.444	0.551	-0.654*	-0.336	0.741***	0.718***
	(0.520)	(0.508)	(0.403)	(0.352)	(0.367)	(0.352)	(0.274)	(0.250)
N	89	86	89	89	221	215	263	264

Notes: Tobit regressions, standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable is the average transfer (0 to 5 EUR) in dictator games (stage 5). *Fairness* reflects fairness perception of the treatment policy, ranging from 1 (completely unfair) to 7 (completely fair). "fw" - "favored winners", "nfw" - "non-favored winners", "fl" - "favored losers", "nfl" - "non-favored losers".

While there is no evidence for retaliation at the aggregate level, it is worth noting that non-favored subjects' transfer less to winners if they perceive an affirmative action policy as less fair (see marginal effects of a tobit regression displayed in Table 2.3: Panel 1). This finding holds both for the group of non-favored subjects overall as well as for the subgroup of non-favored losers. Thus, the absence of evidence on retaliation at the aggregate level masks the fact that winners are retaliated against when affirmative action is perceived as less fair, while they receive higher transfers if it is perceived as fair. However, subjects do not distinguish between favored and non-favored winners.

By pointing out this important role of a policy's fairness perception for its impact on post-competition interactions, our results document a link between previous findings on backlash against favored individuals (e.g., Fallucchi and Quercia, 2018; Leibbrandt et al., 2017) and support for affirmative action policies (Ip et al., 2020).

Result 4 (Post-competition cooperation and retaliation): Overall, the affirmative action policies under study do not affect post-competition cooperation and retaliation. However, winners are retaliated against when affirmative action is perceived as less fair and receive higher transfers if it is perceived as fair.

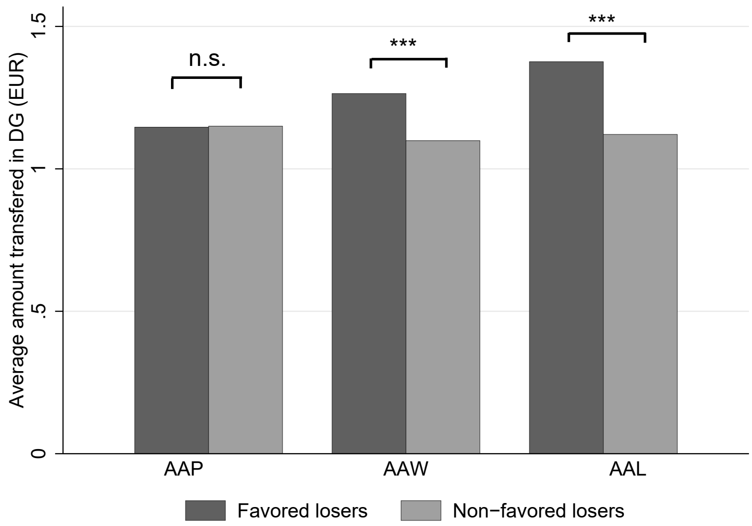

FIGURE 2.6: **Internalization in dictator games**

Notes: Average transfer (EUR) from all group members to favored losers versus non-favored losers in dictator games for each treatment. The brackets and stars above each bar show results of Wilcoxon signed rank tests, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Internalization of affirmative action norms. Finally, our data on post-competition redistributive behavior in stage 5 also allow us to investigate whether subjects "internalize" the normative content of affirmative action policies beyond the context in which they are applied. Intuitively, affirmative action policies aim at improving outcomes for favored individuals. Thus, they convey the social norm that specific individuals should be treated preferentially. In the case of favored losers, this aim has not been achieved. Even though they have been encouraged to compete, they still received the bad outcome (i.e., lost the competition).

In such a situation, subjects who have internalized the norm conveyed by the affirmative action policy could decide to use other available means to support the policy's target group. In the context of the stage 5 dictator games, internalization would imply transferring more to favored than non-favored losers.

As displayed in Figure 2.6, we indeed find evidence in favor of internalization of affirmative action norms that is closely linked to average fairness perception. In treatments AAL and AAW that are rated as significantly fairer than AAP and the control treatment, subjects transfer 23% and 15% more to favored than non-favored losers in the stage 5 dictator games (Wilcoxon signed rank tests: $p < 0.001$ and $p = 0.001$, respectively). This is not the case in treatment AAP (Wilcoxon signed rank test, $p = 0.412$) that is not perceived as fairer than an absence of affirmative action. While we observe that subjects treat favored losers preferentially in exactly those treatments that are generally perceived to be fairer, we do not find a significant relationship with individual perceived fairness of the given treatment (see Table 2.3: Panel 2, right two columns)).

Overall, on average subjects seem to internalize the normative content of those affirmative action policies that are perceived as especially fair and try to act accordingly even beyond the context in which they are binding.

2.4 Conclusion

One defining feature of all affirmative action policies is to base hiring, promotion, admission decisions etc. not on observed performance alone but to complement or adjust observed performance by further, politically desired criteria. For a given technology, ability, effort and luck are the three key determinants of performance (Cappelen et al., 2007, 2010). In this sense, we investigate the "whole universe" of possible quota rules by analyzing the implications of three different quota rules that favor individuals who score low on one of the three determinants of performance. In particular, the quotas favor individuals with low ability (measured by low productivity), with low effort (measured by choosing a short working time), or enduring bad luck, which resembles discrimination, respectively. One advantage of this stylized approach to study the implications of affirmative action policies is that the motivation to favor certain individuals is clearly stated – quite in contrast to a gender quota, for example, that some will attribute to lower skill levels of women, others to offsetting disadvantages due to part-time work, and still others to unjustified discrimination against women. Explicitly stating the reasons for a favored treatment provides a homogeneous perception of what an affirmative action policy is about and a sound basis for eliciting fairness judgments of such policies that can be attributed to the respective criterion for favored treatment.

While Balafoutas et al., 2016; Calsamiglia et al., 2013; Fallucchi and Quercia, 2018; Petters and Schroeder, 2020 study affirmative action policies aiming to compensate for bad luck or discrimination, we are not aware of other attempts to investigate affirmative action favoring individuals with low productivity or short working time, although such policies exist. For example, all countries of the European Union provide part-time employees equal access to promotions even if their overall performance is lower since they work shorter hours. Examples for affirmative action in favor of individuals with low productivity are instances in which individuals with dyslexia or physical restraints get extra time in exams.

Our results document that quotas for discriminated individuals and those who have chosen to work shorter meet their main aim: they effectively encourage the targeted individuals to enter competition. In contrast, a quota for low productivity individuals does not have such an encouragement effect. Compared to a situation without affirmative action, none of the three affirmative action policies under study harms overall efficiency, post-competition teamwork or induces significant retaliation towards the group of favored individuals as a whole. Thus, our results largely reinforce the rather positive findings regarding the consequences of affirmative action policies in studies on gender quotas (e.g., Balafoutas et al., 2016; Balafoutas and Sutter, 2012; Ibanez and Riener, 2018; Kölle, 2017; Niederle et al., 2013) or caste membership (Banerjee et al., 2018, 2020) and extend them to affirmative action policies targeting discriminated individuals and those who perform lower since they have chosen to work shorter hours.

A further result is the high acceptance of affirmative action policies based on judgments regarding their fairness. In particular, affirmative action policies targeting discriminated individuals or those choosing a short working time are judged as significantly fairer than no affirmative action. In times of heated debates about affirmative action, this is important news as it indicates that quotas can get broad political support if they target discriminated individuals or part-time workers and are communicated as such, an aspect that our design ensures.

Perhaps most importantly, our findings suggest that the perceived fairness of affirmative action policies impacts their consequences. Notably, higher fairness perceptions can encourage willingness to compete and prevent retaliation against targeted winners. Additionally, individuals internalize the norms embodied by those affirmative action policies that are rated as fairest and support previously targeted, but unsuccessful individuals even in unrelated post-competition interactions, in which the policy is no longer binding. Affirmative action policies that are perceived as fair thus amplify their impact by influencing behavior beyond the context in which they are applied. From a more general perspective, this finding suggests that not only do preferences shape institutions but that institutions can also shape preferences.

As a whole, our results point at a so far disregarded, but vital aspect in designing and communicating affirmative action schemes, namely their perceived fairness. Providing a convincing rationale for the implementation of quota rules to ensure that they are perceived as fair seems key to make them a success.

2.A Additional results

2.A.1 Variation in productivity and working time

Figure 2.7 and Figure 2.8 show the distributions of productivity and working time for all 463 subjects participating in the preparatory session. The median productivity is 7 in the DICE Lab sample and 8 in the BonnEconLab sample. The median working time is 17 min 46 s in the DICE Lab sample and 17 min 17 s in the BonnEconLab sample. In Figure 2.8, the spike at 61 minutes is due to the fact that we stopped subjects who still worked on the grid task after 60 minutes. Those who work longer are classified as subjects with long working time anyway and any further measurement of their chosen working time is not necessary.

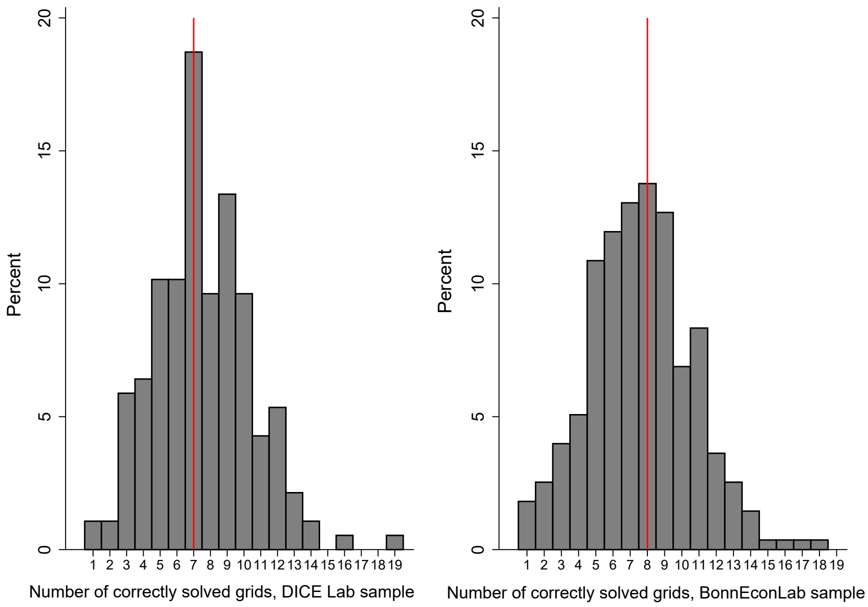

FIGURE 2.7: Distribution and median split (indicated by the vertical red line) of productivity by sample

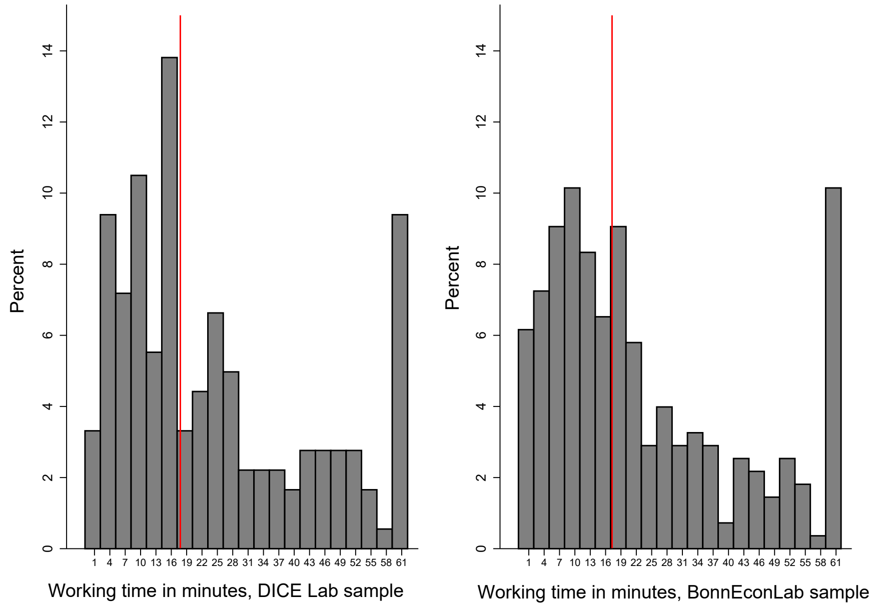

FIGURE 2.8: Distribution and median split (indicated by the vertical red line) of working time by sample

2.A.2 Fairness perception

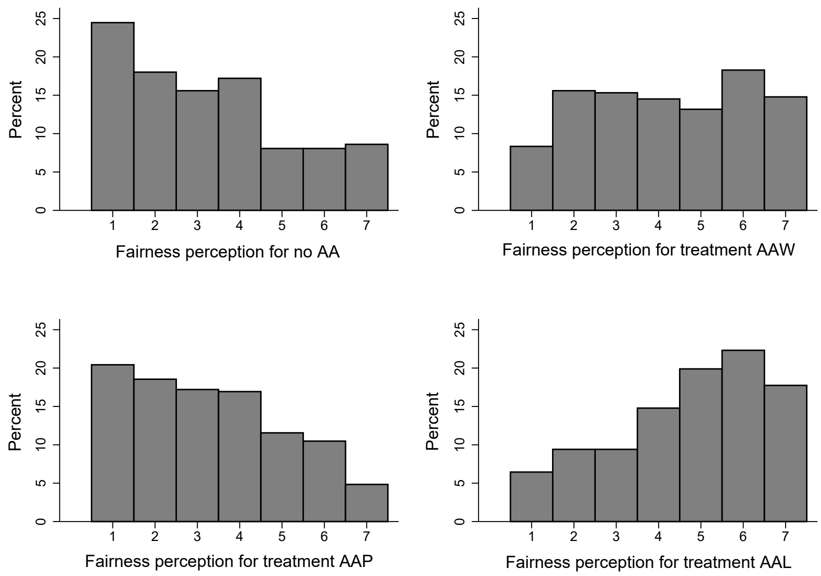

FIGURE 2.9: **Distribution of fairness perception**

Notes: Fairness perception for no affirmative action and for each affirmative action policy. Higher numbers indicate that a policy is perceived as fairer.

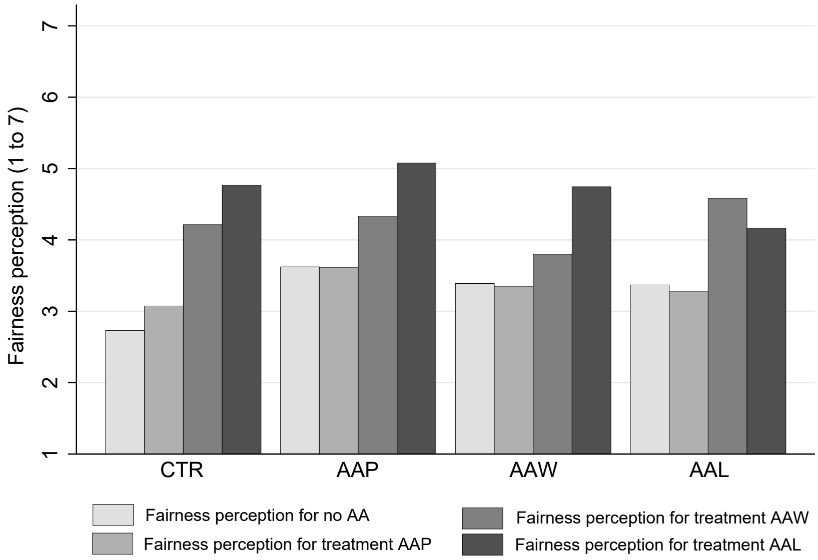

FIGURE 2.10: **Fairness perception by treatment**

Notes: Fairness perception for no affirmative action and for each affirmative action policy by treatment. CTR/AAP/AAW/AAL on the horizontal axis indicate answers from subjects in treatment CTR/AAP/AAW/AAL. Higher numbers indicate that a policy is perceived as fairer.

2.A.3 Willingness to compete

TABLE 2.4: Individual level determinants of competition entry

	(1)	(2)	(3)
Belief of rank			
2	-0.085	-0.057	-0.104
	(0.082)	(0.096)	(0.086)
3	-0.318***	-0.252**	-0.336***
	(0.088)	(0.104)	(0.095)
4	-0.318***	-0.271***	-0.341***
	(0.089)	(0.100)	(0.092)
5	-0.396***	-0.367***	-0.494***
	(0.095)	(0.108)	(0.096)
6	-0.507***	-0.406***	-0.564***
	(0.098)	(0.118)	(0.098)
Risk attitude	0.028***	0.028**	0.031***
	(0.010)	(0.012)	(0.012)
Female	-0.095*	-0.109*	-0.097*
	(0.048)	(0.057)	(0.055)
Fairness perception	0.016	0.028*	0.049**
	(0.013)	(0.015)	(0.021)
Favored			0.496***
			(0.130)
Fairness perception × Favored			-0.069**
			(0.030)
N	372	264	264

Notes: Average marginal effects from a probit regression. Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The binary dependent variable is willingness to compete in stage 3 (1 if tournament, 0 if piece-rate). *Belief on rank* is a categorical variable about beliefs regarding own rank (between 1 and 6) in stage 3; *Risk attitude* is the answer to the general risk question elicited on an 11-point scale, higher numbers indicate a higher willingness to take risks; *Female* is an indicator variable for gender (1 if female, 0 if male); *Fairness perception* reflects fairness rating of own treatment, elicited on a 7-point scale on which higher numbers indicate higher perceived fairness; *Favored* is an indicator variable (1 if favored type, 0 otherwise).

2.A.4 Efficiency decomposition

In this section we take a closer look at the different aspects that enter into the efficiency comparison we conduct in subsection 2.3.3.

The measure used there is tournament winners' number of correctly solved grids per minute. This measure actually captures two separate aspects. The first is winner composition which is determined both by the quota rule (or its absence) and participants' entry decision. The second is how well winners perform under the given incentives.

One thing to keep in mind is that winner selection is not necessarily efficient in the absence of affirmative action. In CTR, the two highest-performing subjects are chosen as winners. Luck and working time are key components of that performance. In fact in CTR in stage 2, only for 28% of groups the winners are the two most efficient subjects, i.e., those who solved most grids per minute correctly in the preparatory session.[25] Theoretically, the quota rule is likely to replace a less efficient individual by a more efficient one as a winner in treatments AAL and AAW, while in AAP this is the other way around. With 36%, 27% and 0% of groups having two most efficient winners in stage 2 in AAL, AAW and AAP, respectively, our data is consistent with this.

We already analyzed the effects of affirmative action on competition entry in section 2.3.2. Although overall effects were modest, we should take a closer look at whether affirmative action encourages tournament entry in stage 3 of the "right" subjects. For this we look at subjects who are among the two most efficient in their respective group according to the number of correctly solved grids in the preparatory session. While in the absence of affirmative action (CTR), 53% of those most efficient subjects enter the tournament, this is the case for 67%, 44%, and 41% of "most efficient" subjects in AAL, AAW, and AAP, respectively. In consequence, 61% of stage 3 tournament winners in CTR are actually the most efficient in their respective group. This share increases to 69% and 63% in AAL and AAW, and decreases to 38% in AAP.

The efficiency measure reported in Figure 2.4 accounts for a possible effect of quota rules on motivation in addition to the effects stated in this subsection.

[25] In this section we regard the two subjects who have the highest number of correctly solved grids per minute in the ability task in the preparatory session as the "most efficient". We use the data from that task since in the preparatory session monetary incentives are the same for everyone. In stage 1 of the main session the assigned type might already affect the number of correctly solved grids per minute through monetary incentives and motivation.

2.A.5 Post-competition measures

TABLE 2.5: Average performance in the slider task

	Overall	Productivity type		Working time type		Luck type	
		High	Low	Long	Short	Lucky	Unlucky
CTR	7.20	7.39	7.02	7.31	7.09	7.24	7.17
AAP	7.08 (0.960)	7.56 (0.586)	6.60 (0.652)				
AAW	6.91 (0.481)			7.21 (0.929)	6.61 (0.255)		
AAL	7.35 (0.115)					7.17 (0.483)	7.52 (0.124)

Notes: Average performance (number of correctly solved slider screens) in stage 4. The first column is average performance by treatment. The following columns show average performance of a specific type in the treatment affecting that type and the control treatment. P-values for Mann-Whitney U tests comparing AA treatments with CTR in parentheses.

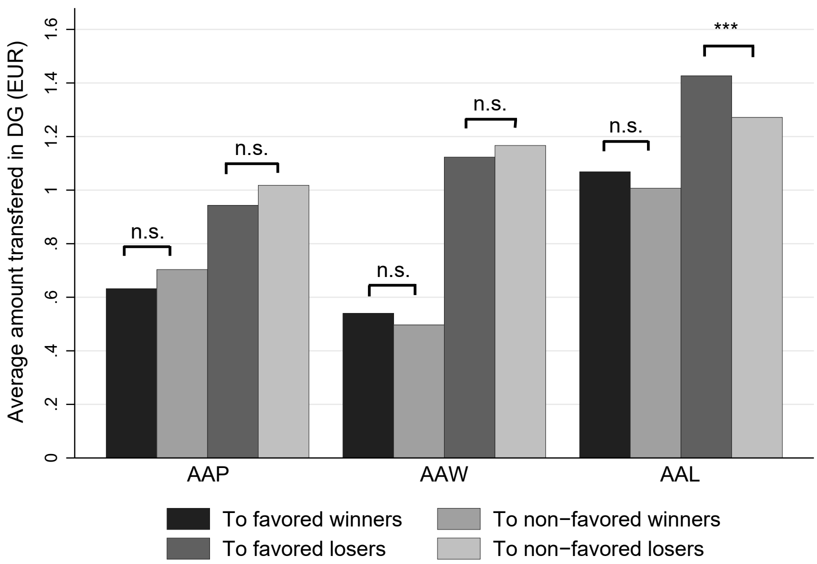

FIGURE 2.11: **Retaliation in dictator games - a broader view**

Notes: The figure displays the average transfer amount in EUR from non-favored subjects to favored winners versus non-favored winners as well as to favored losers versus non-favored losers in the dictator games for each AA treatment. The brackets and stars above each bar show results of Wilcoxon signed rank tests: * $p < 0.1$; ** $p < 0.05$, *** $p < 0.01$.

2.B Measurement of fairness perception

The following text is translated from German and refers to the control treatment (CTR). In all treatments, subjects were first asked how fair they perceive the rules of competition they had actually been exposed to to be, before the other competition rules were described and rated.

In the following, we would like to know how fair you perceived the rules of competition in PART 2 of the experiment to be.

As a reminder, in a group of six members, the two members with the highest overall performance (that is, the number of correct answers × the multiplier) were the two winners of the competition.

How fair do you perceive the rules of the competition to be? The leftmost box means "completely unfair", the rightmost means "completely fair". With the boxes in between you can graduate your statement.

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

Now we would like to know, how fair you perceive other possible rules for the competition to be. Just like in the competition you have participated in in PART 2, the following applies in all competition rules: in a group of 6 members, there are two winners who earn a positive amount of money. The other group members earn nothing. With regard to productivity, working time, and multiplier, the group composition is exactly the same as described on page 2 of the instructions.

Each form of competition has one additional special rule:

- Special rule A: At least one winner must be a group member whose productivity lies in the lower half in comparison to all other participants in the experiment. Productivity is defined as the number of correctly solved tasks per minute of a participant in the five-minute counting task last week.
- Special rule B: At least one winner must be a group member whose working time last week was in the lower half in comparison to all other participants in the experiment, and who therefore worked for 4.5 minutes on the task today.
- Special rule C: At least one winner must be a group member with the low multiplier of 0.75.

How fair do you perceive the competition with special rule A? The leftmost box means "completely unfair", the rightmost means "completely fair". With the boxes in between you can graduate your statement.

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

How fair do you perceive the competition with special rule B?

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

How fair do you perceive the competition with special rule C?

completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

2.C Experimental instructions

(Translated from German)

2.C.1 Preparatory session

The preparatory session started with an on-screen description of the grid task and an unpaid trial round of the grid task (including feedback on whether each table was solved correctly). We then measured baseline productivity, implemented the questionnaire (for details, see section 2.D), and finally measured the individual choice of working time. Below we provide translated versions of the instructions that were originally in German.

Measurement of baseline productivity: On-screen instructions for the five-minute grid task
You will now start working on the task. Your performance in this task is relevant for your payment. The more counting tasks you solve correctly in the given time, the higher your payment. For each correctly solved table, you receive 0.50 EUR. In the upper right corner of the screen, the remaining time (in seconds) is shown. The task lasts 5 minutes.
Please try hard to solve as many tables correctly as possible in the five minutes, so that we get a realistic idea how good you are in this task.

Measurement of choice of working time: On-screen instructions for the grid task in which subjects choose their working time
You will now again work on a similar task as the previous one. As before, you count the number of zeros ("0") in each table and receive 10 Euro-cent for it, i.e. 10 Euro-cent for each correct table. However, you are now free to choose how long you like to work on the task. You will start working on the task on the next

screen, and can work on it as long as you want. The tables will appear one after another, until you decide to stop working.
In addition, there is a special feature: Your working time today will also determine your working time in the experiment next week. Next week, you will work on a similar task again in which you will be given a specific amount of time to solve as many tables as possible and get paid accordingly. In this task, the tables will appear one after another until working time is up. Based on your working time today, we will form two groups of the same size but with different working times. Those who decide to work shorter today, will also work shorter next week (one half of the participants). Those who decide to work longer today, will also work longer next week (the other half of the participants). Individuals who work shorter will, on average, solve fewer tables correctly and therefore earn less. For them, the experiment will be shorter (it will end earlier). Individuals who work longer will, on average, solve more tables correctly and therefore earn more. For them, the experiment will be longer (it will end later). In case you have any questions, please raise your hand. An experimenter will then come to your seat.

2.C.2 Main session

General instructions (distributed on paper at the beginning of the main session)

Welcome to today's experiment! Thank you for participating!

During the experiment, you and the other participants will be asked to make decisions. Your own decisions as well as the decisions of the other participants will determine your earnings, according to the rules that will be described in what follows.
The experiment will be conducted on the computer. You make your decisions on the screen. All your decisions and answers will remain confidential and anonymous.
The experiment consists of 5 parts. PART 1, PART 2, PART 3, PART 4, and PART 5. Additionally, you will answer a short questionnaire.
One of the five parts will be selected randomly by the computer to determine your payment. Every part of the experiment is equally likely to be selected. It is therefore in your own interest to make your decisions in each part as if it was the only part.

Independent of your decisions you will receive a show-up fee of 4 EUR. This means that your total earnings from today's session will be the payment from the randomly chosen one of the five parts of the experiment plus the show-up fee of 4 EUR. You will receive your earnings at the end of today's session together with the earnings from last week.
All other explanations will be given stepwise at the beginning of each part of the experiment. You will receive the instructions for each part in turn. You will have enough time to read the instructions carefully and to ask questions. Please do not hesitate to ask questions if something is unclear.
Please note that, as the last week, talking is not permitted. If you have questions, please do not ask them loudly but raise your hand. One of the experimenters will come to your seat to answer your question. If you do not comply with these rules you will be excluded from the experiment and you will not receive any payments.

General information regarding today's experiment

In today's experiment, your task is once again to solve as many counting tasks correctly as possible in a given amount of time, i.e. to correctly count the number of zeros ("0") in as many tables as possible. In addition, there is one special feature. **Each participant has three characteristics which remain fixed during the whole experiment: his productivity, his working-time (in minutes) and his multiplier**.

- The **productivity** states how many counting tasks per minute the participant has solved correctly in last week's five-minutes-task. For half of the participants, productivity lies **in the lower half**. For the other half of participants, productivity lies **in the upper half**.
- Today's **working time** depends on the self-chosen working time in the task at the end of the session last week. Half of the participants will have **4.5 minutes** per task today to solve as many tables correctly as possible. These are those participants whose working time belonged to the lower half last week. The other half of the participants will have **7.5 minutes** per task today to solve as many tables correctly as possible. These are those participants whose working time belonged to the upper half last week.
- The **multiplier** is a number which is multiplied with the number of correctly solved counting tasks to determine overall performance. The multiplier will be assigned randomly to each participant. For half of the participants, the multiplier will be **0.75**. For the other half, it will be **1.25**.

You will soon receive information about your productivity, your working time and your multiplier on the following screen.
The performance of each participant is determined as follows:

Performance = Number of correctly solved counting tasks in your working time × Multiplier

PART 1 – Piece rate (distributed on paper at the beginning of stage 1)

Your task in PART 1 is similar to the one in the first session. Again, the task is to solve as many counting tasks as possible in a given amount of time, i.e. to correctly count the number of zeros ("0") in as many tables as possible. How much time you have is displayed on the screen. Each table consists of ten rows and ten columns, which contain either a zero ("0") or a one ("1"). Each table differs from the previous one. You are allowed to use the provided scratch paper if you like. After you have entered your response, please click the "confirm" button. Afterwards, you will learn immediately on the same screen whether your answer is right or wrong.
If PART 1 of the experiment is chosen for payment, you will receive the following payment:

$$\textbf{Payment} = \underbrace{\textbf{Correctly solved counting tasks in your working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{0.50 EUR}$$

For example, if you have solved **ten** tables correctly and your multiplier is **1.25**, you receive the following payment:

Payment = 10×1.25×0.50 EUR = 6.25 EUR

If you have answered **ten** questions correctly and your multiplier is **0.75**, you receive the following payment:

Payment = 10×0.75×0.50 EUR = 3.75 EUR

Your payment will not be reduced if you enter a wrong answer. We will refer to this payment as the **piece-rate payment** from now on.
After all questions regarding PART 1 are answered, your working time for PART 1 will start.

PART 2 – Tournament (distributed on paper at the beginning of stage 2)

As in PART 1, you will have a given amount of time to solve as many counting tasks correctly as possible. Again, your working time is displayed on the screen. Different from before, in this part your payment depends on your performance relative to the performance of other participants in your group.
Group allocation:

For the following parts of the experiment, you will be allocated to a **group with 6 members**. The groups were formed randomly and stay the same throughout the whole experiment. This means that you will form a group with the same participants for the rest of the experiment.
Reminder: **Each participant has 3 characteristics: his productivity, his working time, and his multiplier.**
Note that each group consisting of six **members** meets the following criteria regarding productivity, working time and multiplier:

- The **productivity** of **three** group members lies in the **upper half** compared to all participants. The productivity of the **other three** group members lies in the **lower half** compared to all participants.

- The chosen **working time** last week of **three** group members lies in the **upper half** compared to all participants. Therefore, these three group members work for **7.5 minutes** on each counting task today. The chosen **working time** last week of the **other three** group members lies in the **lower half** compared to all participants. Therefore, these three group members work for **4.5 minutes** on each counting task today.

- The randomly drawn **multiplier** of **three** group members is **0.75**. The number of correctly solved tables of these three group members will thus be multiplied with **0.75** to calculate overall performance. The randomly drawn **multiplier** of **the other three** group members is **1.25**. The number of correctly solved tables of these three group members will thus be multiplied with 1.25 to calculate overall performance.

Rules of the tournament:
If PART 2 is chosen for payment, your payment depends on how high your performance is compared to the other five members of your group.
The **two** group members with **highest** overall performance (i.e. number of correctly solved tasks in the total individual working time $\times$ Multiplier) are the two winners of the tournament.
(The content of the following part in gray differs across treatments. There is no further content for the control treatment (CTR).)

Affirmative action w.r.t. productivity (AAP):
In addition, the following **special rule** is applied:
At least one winner must be a group member whose productivity lies in the lower half in comparison to all other participants in the experiment.

Productivity is the number of correctly solved counting tasks per minute last week.
If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose productivity lies in the lower half will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**

Affirmative action w.r.t. working time (AAW):
In addition, the following **special rule** is applied:
At least one winner must be a group member whose working time last week lied in the lower half in comparison to all other participants in the experiment, and who therefore works for 4.5 minutes on the task today.
If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose working time is 4.5 minutes will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members is no longer a winner.**

Affirmative action w.r.t. luck (AAL):
In addition, the following **special rule** is applied:
At least one winner must be a group member with the low multiplier of 0.75.
If this is not automatically the case given the overall performance of the group members, then the group member with the highest performance of the three group members with the low multiplier of 0.75 will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members is no longer a winner.**

The payment of the two winners is as follows:

$$\textbf{Payment} = \underbrace{\textbf{Correctly solved counting tasks in their working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{1.50 EUR}$$

For example, a winner with 10 correct answers and a multiplier of **1.25** receives the following payment:

$$\textbf{Payment = 10} \times \textbf{1.25} \times \textbf{1.50 EUR = 18.75 EUR}$$

A winner with 10 correct answers and a multiplier of **0.75** receives the following payment:

$$\textbf{Payment} = \mathbf{10} \times \mathbf{0.75} \times \mathbf{1.50\ EUR} = \mathbf{11.25\ EUR}$$

The **other four** members of your group get **no payment**.
If there is a tie between two group members, the winner will be determined randomly. We will refer to this payment as **tournament payment** from now on. At the end of today's session, you will be informed about the outcome of the tournament.

PART 3 – Choice between piece-rate and tournament payment (distributed on paper at the beginning of stage 3)

Similar to PART 1 and PART 2, you will have a given amount of time to solve as many counting tasks correctly as possible. Your working time will be shown on the screen.
However, now you choose by yourself which payment scheme you prefer for your performance in PART 3. You can choose either the piece-rate payment (same rules as in PART 1) or the tournament payment (same rules as in PART 2). If PART 3 is chosen for payment, your earnings will be determined as follows:

- If you choose the **piece-rate payment**, your payment is:

$$\textbf{Payment} = \underbrace{\textbf{Correctly solved counting tasks in your working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{0.50 EUR}$$

- If you choose the **tournament payment, your earnings depend on the level of your overall performance in PART 3 compared to the overall performance of your five group members in PART 2 (tournament).** Reminder: PART 2 is the part you have just finished.

(The content of the following part in gray differs across treatments.)

Control treatment (CTR):
If your overall performance (i.e. number of correctly solved counting tasks in the individual working time × Multiplier) is higher than that of at least four other members of your group in PART 2, your payment is as follows:

Affirmative action w.r.t. productivity (AAP):

In general, the two group members with the highest overall performance, i.e. (number of correct answers in the total individual working time) × (Multiplier), are the two winners of the competition.
The following **special rule** is still applied:
At least one winner must be a group member whose productivity lies in the lower half in comparison to all other participants in the experiment. Productivity is the number of correctly solved counting tasks per minute last week.
If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose productivity lies in the lower half compared to all participants in the experiment will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
If your overall performance in PART 3 relative to the overall performance of your group members in PART 2 implies you are a winner, your payment is as follows:

Affirmative action w.r.t. working time (AAW):

In general, the two group members with the highest overall performance, i.e. (number of correct answers in the total individual working time) × (Multiplier), are the two winners of the competition.
The following **special rule** is still applied:
At least one winner must be a group member whose working time last week lied in the lower half in comparison to all other participants in the experiment, and who therefore works for 4.5 minutes on the task today.
If this is not automatically the case given the overall performance of the group members, then the group member with the best performance among the three group members whose working time is **4.5 minutes** will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members is no longer a winner.**
If your overall performance in PART 3 relative to the overall performance of your group members in PART 2 implies you are a winner, your payment is as follows:

Affirmative action w.r.t. luck (AAL):
The following **special rule** still is applied:
At least one winner must be a group member with the low multiplier of 0.75.
If this is not automatically the case given the overall performance of the group members, then the group member with the highest performance of the three group members with the low multiplier of 0.75 will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members is no longer a winner.**
If your overall performance in PART 3 relative to the overall performance of your group members in PART 2 implies you are a winner, your payment is as follows:

$$\textbf{Payment} = \underbrace{\textbf{Correctly solved counting tasks in your working time} \times \textbf{Multiplier}}_{\textbf{Overall performance}} \times \textbf{1.50 EUR}$$

That means it is three times as high as the piece-rate payment.
If your overall performance in PART 3 relative to the overall performance of the other group members in PART 2 implies that **you are not a winner**, you get **no payment**.
If there is a tie between two group members, the winner will be randomly determined.
The group composition is the same as in PART 2. If you choose the tournament payment, you will be informed about the outcome of the tournament at the end of the experiment.
On the next screen, you will decide whether you choose the piece-rate payment or the tournament payment for your performance in PART 3. Then the task will begin.

PART 4 (displayed on screen at the beginning of stage 4)

In the following, you will work on a new task in which you have to place slider markers in a certain position.
You will see six sliders on each screen. They can be placed on a scale from 0 to 100. As soon as you click on a slider marker, the current position will be displayed on the screen. You can change the position using the mouse.
Your task is to move all six slider markers on a screen to the position of "50". Only then a screen is finished correctly and you can proceed to the next screen

by clicking the "Continue" button. You have five minutes to correctly finish as many screens as possible. In this task, all participants work for the same amount of time and there is no multiplier.
Your payment in this part depends on the number of screens that you and the other five members of your group finish correctly. The group composition is the same as before.
Precisely, your payment is determined as follows: You will receive 10 Euro-cent for each correctly finished screen by each member of your group (including yourself). The other members of your group will also receive 10 Euro-cent for each screen that any group member (including yourself) has finished correctly. This means each correctly finished screen by each player yields 60 Euro-cent for the group (i.e. all six group members together).
In addition, the members of the group who won the tournament in PART 2, will receive an endowment of 5 EUR. The other members will receive an endowment of 2 EUR.
If PART 4 is chosen for payment, your payment is the sum of your individual endowment and your earnings from the sum of all correctly finished screens of your group members.
At the end of the experiment, you will be informed about the performance of your group.
If you have any questions, please raise your hand.

PART 5 (displayed on screen at the beginning of stage 5)

In this part, you are asked to make five decisions which will affect you and one of the five other members of your group, respectively. In order to be able to attribute decisions, each group member will be randomly assigned a number from 1 to 6. You are group member number X.[26]
For each decision, you will get an initial endowment of 5 EUR. Your task is to decide how to split this endowment between you and the other member of your group. You may choose an amount between 0 and 5 EUR (in steps of 10 Euro-cents) which you want to pass on to the other group member. You will keep the rest for yourself. You will not get any information about the identity of the other group member and the other group member will not get any information concerning your identity. The only thing you will get to know about the respective other group member before you will make your decision is whether (s)he has won the tournament in PART 2 or not.

[26]The exact number differs for each subject.

(The content of the following part in gray differs across treatments. There is no further content for the control treatment (CTR).)

Affirmative action w.r.t. productivity (AAP):
... and whether his/her productivity lies in the upper or in the lower half.

Affirmative action w.r.t. working time (AAW):
... and his/her working time.

Affirmative action w.r.t. luck (AAL):
... and his/her multiplier.

If this part is chosen for payment, your payment will be determined as follows: In each group, three pairs are chosen randomly and their decisions will determine payments. This means each group member is assigned to exactly one pair which is relevant for the payment. In each pair, it is randomly assigned who will be the donor and the recipient. The decision of the donor determines the payment of both. This means that each decision is paid out with the same probability and therefore you should make each decision as if it was the only one.
At the end of the experiment, you will be informed about the number of the group member you paired with, who is the donor and who is the recipient in this pair, and what the donor has decided. You will not get any information about the decisions made in the other pairs (that you do not belong to).
If you have any questions, please raise your hand.

2.D Questionnaire

The questionnaire in the preparatory session contains the following items (translated from German):

1. **Risk preference, general risk question:** same wording as in German Socio-Economic Panel questionnaire (SOEP, see, for example, Wagner et al., 2007) How do you evaluate yourself? Are you generally a risk-seeking person or do you try to avoid risks? The leftmost box means "not at all risk-seeking" and the rightmost "very risk-seeking". With the boxes in between, you can graduate your statement.

not at all risk-seeking ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ *very risk-seeking*

2. **Risk preference, incentivized choice list:** Subjects make eleven, pairwise decisions between a lottery with a fifty-fifty chance of winning either 2 EUR or 7 EUR and a safe payment. The safe payment increases in 0.5 EUR increments, ranging from 2 EUR to 7 EUR.

3. **Social preference** (survey question, Falk et al., 2018)

   Question 1: Imagine the following situation: Today you unexpectedly received 1000 EUR. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed).

   Question 2: Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize that you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about 20 EUR in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 EUR, the most expensive one costs 30 EUR. Do you give one of the presents to the stranger as a "thank you" gift?

   Which present do you give to the stranger?

   1. No, would not give present
   2. The present worth 5 EUR
   3. The present worth 10 EUR
   4. The present worth 15 EUR
   5. The present worth 20 EUR
   6. The present worth 25 EUR
   7. The present worth 30 EUR

4. **Big Five:** we use the 15-item Big Five scale developed for the SOEP (Schupp and Gerlitz, 2008) to measure personality traits.

5. **Locus of control:** we use 10 different items adapted from Rotter, 1966 which have been used in the 2005 wave of the SOEP.

6. **Questions on general fairness ideals:** all using the same scale

   completely disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely agree*

   To what extent do you personally agree with the following statements?

It is unfair for someone who does a strenuous job to earn little.

Who performs better, should earn more.

If someone is naturally good at something, it is right to reward him/her for it.

It is wrong to favor somebody just because he/she may have experienced discrimination elsewhere.

7. **Cognitive ability:** Raven matrices from the Wechsler IQ test (Raven and Raven, 2008).

   Before subjects start the test, we elicit their belief about individual rank as follows:

   Before you begin, we would like to ask you to assess how well you will score in the IQ test compared to the other participants in the experiment. For example, 0-10% means that you are among the 0-10% participants with the fewest correct answers, and at least 90% of the participants have more correct answers than you. 41-50% means that at least 40% of participants have fewer correct answers and at least 50% have more correct answers than you. 91-100% means that at least 90% have fewer correct answers than you. What do you think? How do you compare to the rest of the group?

   ☐ 0-10%
   ☐ 11-20%
   ☐ 21-30%
   ☐ 31-40%
   ☐ 41-50%
   ☐ 51-60%
   ☐ 61-70%
   ☐ 71-80%
   ☐ 81-90%
   ☐ 91-100%

8. **Cognitive reflection test:** see Frederick, 2005.

9. **Socio-demographics:** age, gender, final grade point average at academic high school, last math grade at academic high school, field of study, monthly disposable amount of money, political orientation, number of experiments already participated in the same lab.

Declaration of contribution

Hereby I, Chi Trieu, declare that this chapter, entitled "Fairness perception and consequences of affirmative action policies" is co-authored with Hannah Schildberg-Hörisch and Jana Willrodt. The contributions of each author to this paper is as follows:

- The conception of the initial idea was done by Hannah Schildberg-Hörisch.
- All authors jointly and equally developed the specific concepts of the paper and the experimental design.
- Programming the Z-tree codes was done by Chi Trieu, with substantial support from Jana Willrodt.
- Conducting the experiment was done by Jana Willrodt and Chi Trieu.
- All authors jointly and equally contributed to the empirical analysis and graphical representation of the results.
- All authors jointly and equally contributed to the writing of the manuscript.

Signatures of the co-authors:

Hannah Schildberg-Hörisch | Jana Willrodt

Chapter 3

How uncertainty about the favored group affects outcomes of affirmative action

3.1 Introduction

Affirmative action is implemented by institutions around the globe.[1] Nevertheless, it remains a highly controversial policy, even among its proponents. One critical opinion points to the potential over- and under-inclusiveness of conventionally practiced affirmative action policies targeting one specific disadvantaged group. Notably, favoring one particular minority seems subjective and might fail to fully capture the truly disadvantaged individuals in an increasingly diverse population and workforce.[2] Advocates of this viewpoint call for the coexistence of affirmative action policies targeting several marginalized groups in order to better capture structural determinants of disadvantage.

Many institutions have reformulated their diversity management policies accordingly. For example, the European Union recognized the importance of anti-discrimination measures not only for gender and ethnic minorities, but also regarding age, disability, religion, and sexual orientation (Hyman et al., 2012), thus mandating member states to address discrimination in these areas in employment and occupation (Council of European Union, 2000). In Germany, since 2001 employers with more than 20 employees have had to reserve at least 5% of positions for individuals with disabilities (SGB IX – *German Social Code, Volume 9*, 2001) and since 2016, at least 30% of the supervisory or administrative boards of publicly listed companies have to be filled by the underrepresented gender (Federal Law Gazette, 2015). In India, seats at higher education institutions are reserved through two quota systems, one for women and one for members of certain disadvantaged social groups (castes) (Bagde et al., 2016).

This paper aims to experimentally investigate one key feature of the simultaneous occurrence of several affirmative action policies, namely the uncertainty about the targeted groups. Clearly, firms, governments, and universities as decision-makers know how affirmative action policies influence outcomes of hiring, promotion, or admission decisions. However, this information is usually not publicly communicated. Hence, from the applicants' side, there is an inherent uncertainty whether or not they were actually favored, even if there does exist an affirmative action policy explicitly targeting them. For example, if

[1]See, e.g., Holzer and Neumark, 2000 for affirmative action policies in the U.S., Sowell, 2004 for affirmative action policies in India, Malaysia, Nigeria, and Sri Lanka, and Hyman et al., 2012 for diversity management tools, including affirmative action policies under the legislation of the European Union.

[2]Take gender quotas for example, several studies reveal the potential inefficiency when classical separation of advantaged and disadvantaged groups are applied. For example, men also shy away from tournaments in team competition (Dargnies, 2012). Apart from gender, other factors also influence willingness to compete, such as socio-economic backgrounds (Almås et al., 2016), age, context, and type of task (Dreber et al., 2011, 2014).

both gender quotas and quotas for individual with disabilities exist, female applicants are uncertain of whether they are in fact favored or not. Consequently, the identity of the (non)favored group is not salient. We build on this key feature and examine how it influences multiple outcomes of affirmative action policies in the form of quota rules in a laboratory experiment.

Affirmative action is typically studied in a competitive setting in order to resemble real-world admission, hiring, and promotion processes. In such settings, the uncertainty and salience of group identity play a role in how individuals and groups respond to competitive incentives (see section 3.2). Despite the emerging empirical literature on the effectiveness and consequences of affirmative action policies, much less is known about the relationship between these factors and outcomes of affirmative action, which is the main focus of this study.

In a competitive setting, we form group identities by characterizing each subject by gender (man or woman) and a randomly given color (green or blue). Both gender and the stylized color characteristic allow us to create completely distinct groups, which is a useful feature for our analysis.[3] Our main treatment variation is the occurrence and criteria of affirmative action policies. In the baseline treatment, no quota is implemented and thus no specific group identity is enhanced. In a second treatment, a 50% gender quota is applied to target women, making the identity of the favored group salient and fully observable. In a third treatment, either women or members of one color group are favored by a 50% quota rule with a fifty-fifty chance. As a result, it is uncertain whether women is the favored group, and the identity of the (non)favored group is not salient.

Except for affirmative action targeting gender (which is one of our treatments), there is generally a fifty-fifty percent gender ratio in disadvantaged groups, such as students with a low socioeconomic status or individuals with disabilities. Hence, we set an equal probability for men and women to belong to one or the other color group. In addition, gender is observable while colors are unobservable to others. Our assignment of group identities is crucial for causal identification, yet it is surely stylized compared to competition in the labor market or admission to higher education. In section 3.5, we discuss the implications as well as the possible limitations of our approach.

We examine both immediate and spillover effects under different policies. First, we test the immediate outcomes by comparing efficiency in terms of task performance and how candidates are selected, as well as willingness to compete in tournaments across treatments. Second, we investigate spillover outcomes in

[3] It is noteworthy that artificially inducing identity does not affect the degree of discrimination in lab experiments compared to using natural identities (see, e.g., Lane, 2016 for a meta-analysis).

a following teamwork setting. Subjects first have to select two out of six group members to form a team. They then work on a new real-effort task under team incentives, meaning the total team performance is equally shared regardless of each team member's contribution. In this setting, we measure perceived competence of (un)favored group members via team choice, and individual effort provision in a team. While building on the classical design of Niederle and Vesterlund, 2007, our experiment contributes in two ways: (i) we introduce uncertainty about the favored group, thus varying the salience of identity of the beneficiaries across treatments, and (ii) we add perceived competence as a new measure of the spillover outcomes of affirmative action.

Regarding the second contribution, our motivation to measure perceived competence is twofold. First, a fair amount of studies show that performance in real-effort tasks with a short working time implemented in laboratory experiments is difficult to influence. In such settings, subjects usually exert maximal effort.[4] This tendency might mask the null result that affirmative action does not harm cooperation observed in previous studies (Balafoutas et al., 2016; Kölle, 2017) in which cooperation is measured by effort provision. Our design introduces additional measures of cooperation that are not subject to such constraints. Second, team selection most likely occurs in real-life workplaces, where many projects allow employees to self-select into teams with colleagues. In this scenario, a comprehensive measure of spillover effects of affirmative action is not only effort provision but also the structure of teams and the efficiency gain (loss) in terms of team performance, if the team members had been selected differently.

We find that uncertainty about the favored group does not affect efficiency measured by performance. There are no significant differences in performance, neither of subjects selecting competition nor of the subgroup of winners across treatments. When affirmative action explicitly targets only gender, it does not encourage women to enter competition. In contrast, when both women and one color group are targeted, competition entry of the favored groups significantly increases. Controlling for risk preference, we propose that this effect operates through beliefs. In particular, women under gender quotas are significantly less confident about their relative rank compared to women and favored types in other treatments. The difference in belief is not attributed to an actual difference in performance, suggesting that gender quotas lead women to self-stigmatize their competence, hence lowering their willingness to compete.

[4]See, e.g, Corgnet et al., 2015, Araujo et al., 2016, Gächter et al., 2016, Goerg et al., 2019.

In the post-competition setting, we find a general misperception of competence toward women in the baseline treatment. They are selected significantly less for teamwork, although their performance is not lower than men's. More strikingly, this gap becomes wider under gender quotas, but narrower and insignificant when women are not favored with certainty.

Both immediate and post-competition outcomes of different quotas lead us to conclude that the salience of identity of the favored group plays a role in determining the consequences of affirmative action policies. When there is one favored group and its identity is easy to make salient, affirmed individuals might be perceived as less competent, both by themselves and by others. This is not the case when the identity of the favored groups is uncertain. Our findings imply that negative consequences of gender quotas observed in previous studies with women as the solely favored group (Ip et al., 2020; Leibbrandt et al., 2017) might well describe the "worst-case scenario". Furthermore, simultaneously targeting several disadvantaged groups and avoiding priming group identities as the (non)favored play an important role in making affirmative action work.

The rest of the paper proceeds as follows. Section 3.2 puts our study into perspective with the related literature, section 3.3 explains the experimental design, section 3.4 shows the experimental results, and section 3.5 discusses the implications of our main results and concludes.

3.2 Related literature

Our research is directly related to a growing line of empirical literature evaluating consequences of affirmative action policies. A substantial amount of literature along the same line, studies quotas for women, racial and ethnic minorities, resembling real-world policies targeting these groups.[5] Apart from quotas, other forms of studied affirmative action include bonuses and lump sum payments (Balafoutas and Sutter, 2012; Calsamiglia et al., 2013; Schotter and Weigelt, 1992), repetition of competition conditional on a threshold of positions being filled by the favored group (Balafoutas and Sutter, 2012), reservation of an extra prize for the targeted group (Fallucchi and Quercia, 2018), and statements

[5]Gender quotas have been examined in numerous settings such as laboratory experiments (Balafoutas et al., 2016; Balafoutas and Sutter, 2012; Beaurain and Masclet, 2016; Ip et al., 2020; Kölle, 2017; Leibbrandt et al., 2017; Maggian and Montinari, 2017; Niederle et al., 2013), field experiments (Beaman et al., 2009; Chattopadhyay and Duflo, 2004; Ibanez and Riener, 2018), and empirical studies on mandatory quotas on corporate boards in Norway (Ahern and Dittmar, 2012; Bertrand et al., 2018; Matsa and Miller, 2013). Quotas targeting castes in India (Bagde et al., 2016; Banerjee et al., 2018; Jensenius, 2015; Pande, 2003) and students of ethnic minorities in the U.S. (see, e.g, Arcidiacono and Lovenheim, 2016) have also been extensively researched.

of equal employment opportunities (Leibbrandt and List, 2018). We contribute by investigating outcomes of affirmative action policies when there are two targeted groups, and the affirmed group is uncertain.[6] As aforementioned, there are several examples as such in reality.

More broadly, our study bridges the affirmative action literature with the literature on group identity and uncertainty in competitive settings. A more salient group identity is linked to changes in individual willingness to compete (Cornaglia et al., 2019; Gupta et al., 2013), more intense competition between groups (Chen et al., 2015; Kato and Shu, 2016), and increasing group conflicts (Chowdhury et al., 2016). Balafoutas and Sutter, 2019 show that uncertainty or ambiguity in number of winners strongly shifts the gender differences in performance and willingness to compete in favor of male participants. Gee, 2019 finds that simply removing the uncertainty about the number of applicants increases application finish rate, and encourages women and minorities to apply for jobs.

Outcomes of affirmative action policies documented in experimental studies are by far ambiguous. On one hand, gender quotas increase competition entry of women, at the same time do not distort men's willingness to compete and do not harm efficiency (Balafoutas and Sutter, 2012, Niederle et al., 2013, Balafoutas et al., 2016. Beaurain and Masclet, 2016 find that gender quotas improve subjective evaluation of women's competence in a hiring game, thus increasing their employment. Quotas for castes enhance confidence and thereby raise willingness to compete of the favored group (Banerjee et al., 2020). Regarding spillover effects, gender quotas do not result in subsequent unethical behaviors (Banerjee et al., 2018; Maggian and Montinari, 2017) or spiteful behavior against the affirmed group (Banerjee et al., 2018). Neither is this policy harmful to effort provision in teamwork (Balafoutas et al., 2016, Kölle, 2017), willingness to work in team (Kölle, 2017), or coordination with team members (Balafoutas and Sutter, 2012).

On the other hand, gender quotas might backfire and generate sabotage

[6] Bagde et al., 2016 and Leibbrandt and List, 2018 study affirmative action policies for more than one group of beneficiaries. Bagde et al., 2016 study quotas for castes and women in university admission using data from India. Outcomes of quotas are evaluated through college attendance and the quality of selected colleges of the targeted groups. Leibbrandt and List, 2018 investigate the impact of a statement emphasizing equal opportunity regarding gender, racial minorities, age, and other protected characteristics on job entry. Compared to these studies, our research differs significantly in the outcomes investigated, and uncertainty and salience of group identity as the main aspect of interest.

against women, consequently discouraging them from entering competition (Leibbrandt et al., 2017). Affirmative action in the form of an equal employment opportunity statement is found to discourage rather than empower the racial minority in job applications (Leibbrandt and List, 2018). When the favored group is randomly assigned, retaliation threat is detrimental to the competition entry of the favored group (Fallucchi and Quercia, 2018), and a quota reduces cooperation in following public good games (Mollerstrom, 2019).

The pros and cons of affirmative action pave the way for a new strand of studies emphasizing the importance of providing evidence of discrimination against the targeted group (Ip et al., 2020), justifications (Petters and Schroeder, 2020), and fairness perceptions (Schildberg-Hörisch et al., 2020) in guaranteeing positive outcomes of affirmative action policies. Our study supplements research in this vein by adding the perspective of uncertainty and the salience of group identity, and shows its impact on consequences of affirmative action in the form of quota rules.

3.3 Experimental Design

3.3.1 The real-effort task

We apply affirmative action on the performance of an arithmetic task introduced by Niederle and Vesterlund, 2007 in which subjects have to add as many sets of five 2-digit numbers as possible in a given time. No calculator is allowed, but subjects can use pen and scratch paper while working on the task. After each attempt, subjects move directly to the next questions without getting feedback whether or not their answer was correct.

Measuring performance by the arithmetic task serves two purposes. First, no special knowledge is required, and a minimal learning effect has been documented in previous studies (Balafoutas and Sutter, 2012; Niederle and Vesterlund, 2007). Second and more importantly, previous studies show that gender difference in willingness to compete is sensitive to the nature of the tasks. In particular, women are less competitive in mathematical tasks than men (see, e.g., Sutter et al., 2019). Thus, using this task to measure performance, we introduce a reasonable need for gender quotas in our setup.

3.3.2 Stages and treatment design

The experiment consists of five stages and a final questionnaire. Each stage is designed as follows:

Stage 1: Piece-rate. Subjects work on the task for five minutes and receive a piece-rate payment of 0.5 EUR per correct answer. Before starting stage 1, we give subjects a 2-minute trial round to eliminate the learning effect. The purpose of stage 1 is to measure the baseline performance without competitive incentives.

Stage 2: Tournament. In stage 2, subjects work under tournament incentives. Each subject competes in a group of six, with three men and three women, three Blue types and three Green types. Each subject has a 50% chance of having either color, independent of gender.Subjects stay in the same group throughout the experiment. Two winners are selected in each group. Each winner earns 1.5 EUR per correct answer, while the losers earn nothing. A random tie-breaking rule is applied in the case of a tie. The purpose of this stage is to measure the effect of different affirmative action schemes on performance and the composition of winners. Subjects are informed about their types at the beginning of stage 1, and group structure is informed at the beginning of stage 2. Subjects never learn the type of other subjects. To avoid any wealth effects, we inform results of the tournament at the end of the experiment.

We implement a *between-subject treatment design* which determines how the winners are selected.

1. **Baseline treatment (CTR)**: winners are the two best performers.
2. **Gender quotas treatment (GQ)**: there must be at least one woman among the two winners.
3. **Mixed quotas treatment (MIX)**: gender quota is applied with a 50% chance, and color quota is applied with another 50% chance. For gender quotas, there must be at least one woman among the two winners. For color quotas, there must be at least one Green type among the two winners.

The treatment design implies that in all treatments, the best performer is always one of the two winners. If the quota rules are not fulfilled, the second-best winner is replaced by the best performer from the favored groups.[7]

[7] We make sure that subjects understand the rule of the competition via several control questions. Each control question is about a made-up competition with information on performance and the favored group. Subjects need to correctly select the winners in all control questions before they can start working on the task.

Stage 3: Choice between piece-rate and tournament. After subjects have experienced both piece-rate and tournament incentives, we ask them to choose which scheme they want to apply to their score in this stage. If the tournament incentive is chosen, their score is compared to stage 2 score of the other five group members, irrespective of their choices (Niederle and Vesterlund, 2007). Winners are then selected with the same rules and payoffs as in stage 2. If the piece-rate is chosen, subjects receive 0.5 EUR per correct answer. Stage 3 is to elicit the willingness to compete.

At the end of stages 1, 2, and 3, subjects receive feedback on their own performance of the respective stage. Subjects never know about the performance and choices of other subjects.

Stage 4: Submit stage 1 score to either piece-rate or tournament. We ask subjects to choose their preferred scheme to apply for their score in stage 1. If the tournament is chosen, their stage 1 score is compared with the stage 1 scores of the other five group members, irrespective of their choices. The winners are then selected with the same rule and payoffs as in stage 2. If the piece-rate is chosen, subjects receive 0.5 EUR per correct answer in stage 1. Asking subjects to submit their past performance to their preferred scheme accounts for preference for performing on top of preference for competing (Ifcher and Zarghamee, 2016). The outcome of stage 4 is independent and does not replace the outcome of stage 1.

Belief elicitation: At the end of stage 4, we elicit beliefs about relative performance in stage 1, stage 2, and stage 3 both within the group of six and within the group of same gender/color. We incentivize by randomly selecting one guess to be paid. Subjects receive 1 EUR if their selected guess is correct (see section 3.B).

Stage 5: Partner selection under team incentives. Subjects move to a new working setup with a new task (Grid task – Abeler et al., 2011), though they remain in the same group and the same treatment.[8] In this stage, subjects are first asked to build their own team of three by selecting two among five group members. Each team of three must include one leader and two colleagues. Leaders (colleagues) must be one of two winners (two of four losers) in stage 2. When selecting team members, subjects are informed about the gender of each group member, and whether they are leaders or not. Naming winners (losers) in stage

[8]In this task, subjects count the number of zeros in a 10-by-10 table containing 100 digits of randomly distributed zeros and ones. Similar to the arithmetic task, this task does not require special prior knowledge or skills.

2 as leaders (colleagues) is to emphasize the hierarchical order from stage 2 tournament outcome. For the same purpose, we introduce unequal bonuses for leaders and non-leaders (similar to Balafoutas et al., 2016). Subjects receive a bonus of 5 EUR if they are the leaders, and 2 EUR otherwise.

Subjects then work on the Grid task for five minutes under a team incentive. The overall payoff of the team is equally shared among all members. Each correctly solved grid generates 0.3 EUR for the team, i.e., 0.1 EUR for each member.

Subjects do not learn the partner selection of other group members. One subject's choice of partners does not affect the partner choice of others, such that partner selection is an individual decision and does not carry externalities on group members. By making partner selection an individual choice, we rule out the redistributive motive and the incentive to sabotage.[9] We assume that subjects have standard preferences, and make a team choice in order to maximize their payoffs in this stage. Optimally, each subject should opt for two group members who they believe can contribute the most to their team. For a leader, the optimal team selection is to choose the two best performers out of four non-leaders. For a non-leader, the optimal team formation is to choose the better performer out of two leaders, and the best performer out of three non-leaders. Hence, decisions on partner selection directly reflect perception about the competence of other group members in the Grid task.[10]

At the end of the experiment, subjects answer a questionnaire which includes our control variables. The questionnaire contains measures on risk and social preferences, positive and negative reciprocity, fairness perception, cognitive ability, overconfidence, and socio-demographics (see section 3.D). Table 3.1 summarizes our experimental design.

3.3.3 Procedures

The experiment is registered in the AEA RCT Registry.[11] Overall, we adhered to the registered number of observations as well as the experimental design. We conducted the experiment at the DICE Lab at the University of Düsseldorf in December 2019 and January 2020 using the software zTree (Fischbacher, 2007).

[9]E.g., a high performer selects into a team with a low performer, or a subject selects into a team with another group member, then deliberately perform worse to decrease the group payoff at own cost.

[10]Performance in the Arithmetic task and in the Grid task are moderately correlated (Pearson's correlation coefficient: $\rho = 0.412$).

[11]Chi Trieu, 2019. "Identity of affirmed groups and the consequences of affirmative action policies" December 02. Registry number AEARCTR-0005122.

TABLE 3.1: Summary of experimental design

Stage	
	Practice round (two minutes, arithmetic task)
Stage 1	Piece-rate incentive (five minutes, arithmetic task)
Stage 2	Competition incentive (five minutes, arithmetic task)
Stage 3	Choice between piece-rate incentive and competition incentive (five minutes, arithmetic task)
Stage 4	Choice between piece-rate incentive and competition incentive for stage 1 performance (no task)
Stage 5	Team selection under team incentive (five minutes, grid task)
	Questionnaire

Subjects were recruited via ORSEE (Greiner, 2004) from the subject pool of the DICE Lab.

In total, 240 subjects of various disciplines participated in the experiment, with 64, 84, 90 subjects in CTR, GQ, and MIX, respectively. One out of the five stages of the experiment is randomly selected to be payoff-relevant. The total payoff consists of a 4 EUR show-up fee, the earnings from the selected stage, and a 2 EUR fixed payment for answering the questionnaire. On average, each subject earned 13 EUR, and each session lasted 90 minutes.

3.4 Experimental Results

In this section, we first examine efficiency in terms of task performance. We then study the willingness to compete at the aggregate level. Thereafter, we characterize the determinants of willingness to compete at the individual level and provide evidence that gender quotas lead to self-stigmatization. Finally, we study perceived competence in team selection, and team performance as post-competition outcomes.

3.4.1 Efficiency

One of the main arguments from proponents of affirmative action is that this policy does not lower the quality of either candidates or winners in competitions, and thereby does not harm efficiency. In our setup, we evaluate efficiency in terms of task performance (number of correct answers in the arithmetic task) of the candidate pool and of the selected candidates. In particular, we compare performance of winners in the compulsory competition in stage 2, subjects who choose tournament incentive over piece-rate incentive, and the subgroup of winners in the self-selected competition in stage 3.

In the compulsory competition, we observe no significant differences in the task performance of winners across treatments. In CTR, winners solve an average of 13.68 tasks correctly, while this number in GQ is 14.32 and in MIX it is 14.5 (two-sided Mann-Whitney U test, CTR vs. GQ: p= 0.568, CTR vs. MIX: p= 0.356, and GQ vs. MIX: p= 0.638).[12] A Kruskal-Wallis test does not detect any significant differences across all treatments (p = 0.631).

In the self-selected competition, Figure 3.1 shows the task performance in stage 1 and stage 3 of subjects who choose tournament incentive in stage 3 in the left panel, and of winners in the right panel. Again, we do not observe any significant differences. The null hypotheses on the equality of means cannot be rejected for any pairwise comparisons of treatments (test results reported in the figure note).[13]

Result 1: Both gender quotas and mixed quotas do not cause efficiency loss compared to no quotas.

Result 1 reinforces a desirable feature of gender quotas that this policy does not affect efficiency measured by how winners are selected and their performance (Balafoutas et al., 2016; Balafoutas and Sutter, 2012; Niederle and Vesterlund, 2007). It further shows that this feature also holds true when quotas target two groups. However, it is noteworthy that in our setup, mixed quotas are implemented without the threat of sabotage or retaliation against affirmed individuals. With retaliation, Fallucchi and Quercia, 2018 find that efficiency in terms of performance is hampered compared to without retaliation.

In addition, our data provide evidence of the efficiency-enhancing effect of competition (Balafoutas et al., 2016; Niederle and Vesterlund, 2007 and Figure 3.4). In all treatments, performance increases significantly when moving from piece-rate incentive to tournament incentive (Wilcoxon signed-rank test, $p < 0.001$ for all treatments). Under both incentives, the average performance of women is not significantly lower than men.[14]

[12] Throughout the paper, unless noted otherwise, we report the results of two-sided tests.

[13] Extending the analysis to performance of all subjects in stage 1, we obtain a similar result that performance does not significantly differ across treatments, confirming the validity of our sampling (see Figure 3.4).

[14] In stage 1, women solve an average of 9.04 tasks correctly, while men score 10.08. In stage 2, the performance of women increases to 10.02 and the performance of men rises to 11.19 (Mann-Whitney U test, p= 0.128, p= 0.253, p= 0.939 for performance in stage 1, p= 0.238, p= 0.105, p= 0.997 for performance in stage 2 for CTR, GQ and MIX respectively).

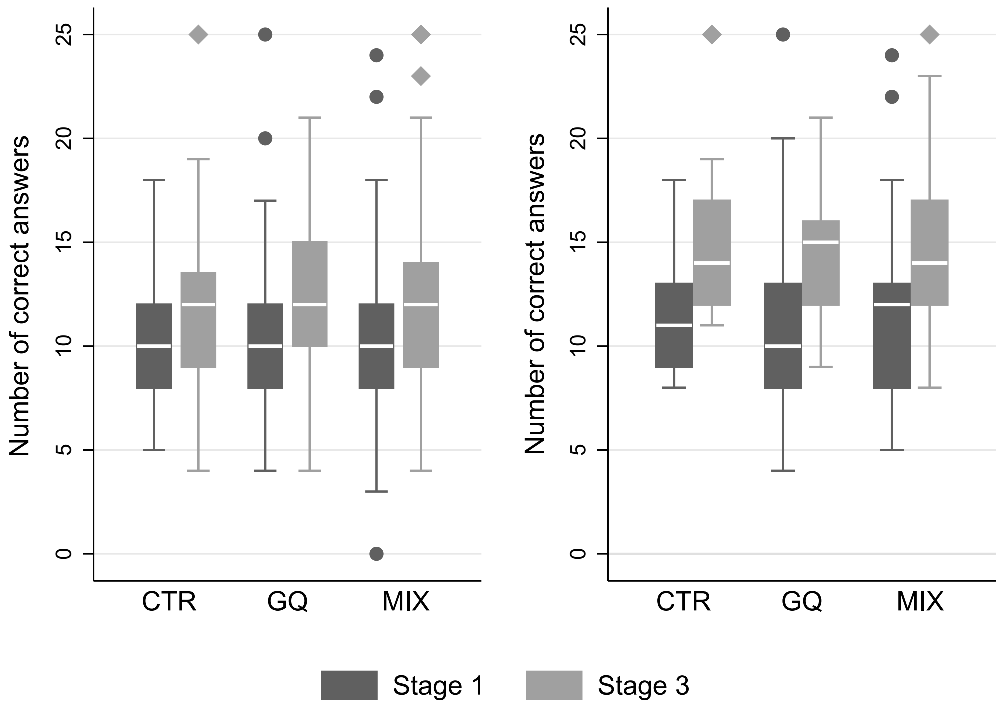

FIGURE 3.1: **Performance in stage 1 and stage 3 of subjects self-selecting into competition and of winners in stage 3.**

Notes: The figure displays performance under piece-rate and tournament incentives of subjects who self-select into tournament (left panel) and of winners (right panel) in stage 3. The upper (lower) hinges of the boxes show the 75th (25th) percentiles, the white lines inside the boxes show the median values, the upper (lower) adjacent lines show the maximum (minimum), and the points show outliers. No significant difference in efficiency across treatments. The average performance of subjects opting for tournament in stage 1 (stage 3) in CTR is 10.22 (11.69), in GQ is 10.54 (12.16) and in MIX is 10.24 (12.19). Mann-Whitney U test, CTR vs. GQ: p= 0.903 (p= 0.568), CTR vs. MIX: p= 0.885 (p= 0.710), GQ vs. MIX: p= 0.782 (p= 0.827) for performance in stage 1 (stage 3). Kruskal-Wallis tests, p= 0.962 (p=0.852) for performance in stage 1 (stage 3). The average performance of stage 3 winners in stage 1 (stage 3) in CTR is 11.44 (14.88), in GQ is 11.30 (14.35) and in MIX is 11.77 (14.46). Mann-Whitney U test, CTR vs. GQ: p= 0.597 (0.761), CTR vs. MIX: p= 0.907 (p= 0.715), GQ vs. MIX: p= 0.547 (p= 0.894) for performance in stage 1 (stage 3). Kruskal-Wallis tests, p= 0.803 (p=0.925) for performance in stage 1 (stage 3).

3.4.2 Selection into tournament

Willingness to compete at the aggregate level

The left panel of Figure 3.2 shows selection into competition for each gender in each treatment. We observe an encouraging effect of quotas on the willingness to compete of women in MIX, but not in GQ. Under gender quotas, 42.9% of women choose tournament, only 0.5 percentage points higher than the competition entry rate of women in CTR (Fisher's exact test, p= 1.000). Under mixed quotas, 66.7% of women select into tournament, 24.3 percentage points significantly higher than in CTR (Fisher's exact test, p= 0.040) and as high as the willingness to compete of men in CTR.[15]

[15]Men opt for competition more than women per se. In CTR, 66.7% (57.6%) of men choose tournament in stage 3 (stage 4), significantly more than the proportion of women at 42.4%

As can be expected, there is an discouraging effect of quotas on the competition entry of men. In GQ, 45.2% of men choose tournament, 21.5 percentage points lower than the competition entry rate of men in CTR (Fisher's exact test, p= 0.101). Men in MIX are also discouraged from selecting the competitive incentives compared to CTR, but less so compared to GQ. Their competition entry rate is 53.3%, 13.4 percentage points lower than in CTR (Fisher's exact test, p= 0.255) and 8.1 percentage points higher than men in GQ (Fisher's exact test, p= 0.522).

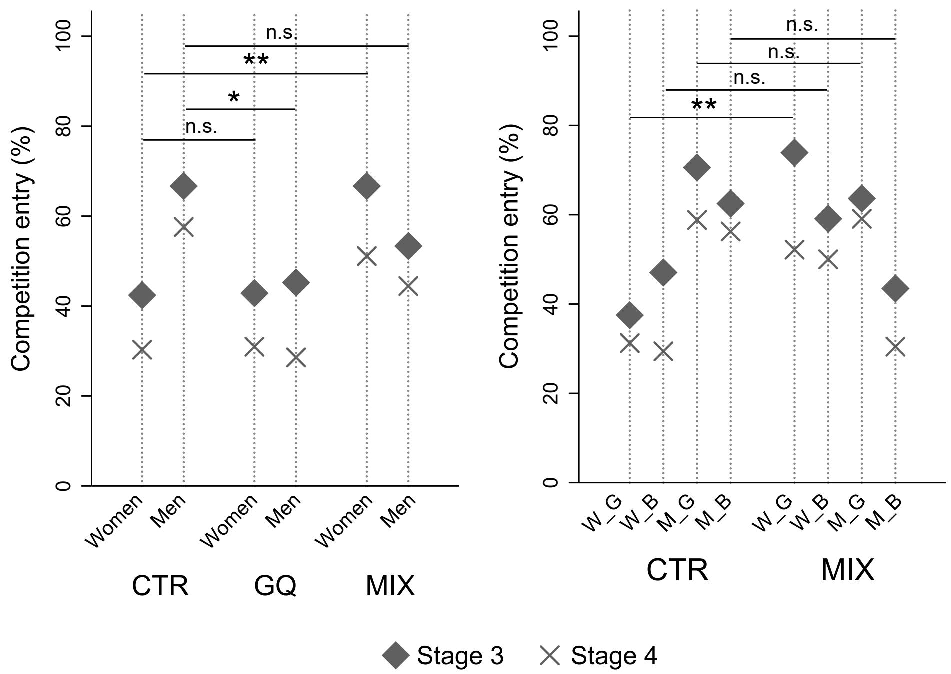

FIGURE 3.2: **Willingness to compete by treatment, gender and type**

Notes: The left panel displays the proportion of competition entry in stage 3 and stage 4 for each gender in each treatment. The right panel displays the proportion of competition entry in stage 3 and stage 4 for four types in treatment CTR and treatment MIX. The brackets and stars above each line show the results of Fisher's exact tests for competition entry in stage 3, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The absence of stars shows that a difference is not significant. Abbreviations: W_G - Women_Green, W_B - Women_Blue, M_G - Men_Green, M_B - Men_Blue. Regarding the right panel, 73.91% of W_G choose the tournament incentive in stage 3, higher than the same type in CTR at 37.5% (Fisher's exact test, p=0.046). 59.09% of W_B in MIX enter competition, in comparison with 47.06% in CTR (Fisher's exact test, p=0.528). Competition entry of M_G in MIX is 63.64% compared to 70.59% in CTR (Fisher's exact test, p=0.740). 43.48% of M_B in MIX enter competition, 19.02 percentage points lower than the same type in CTR (Fisher's exact test, p=0.333).

The right panel of Figure 3.2 shows competition entry by types in treatment CTR and treatment MIX. In MIX, the quotas favor women with a 50% chance,

(30.3%)(Fisher's exact test, p= 0.041 (0.023)). This evidence is in line with the findings of previous studies about the gender gap in the willingness to compete (Gneezy et al., 2003, Niederle and Vesterlund, 2007, Croson and Gneezy, 2009, Leonard et al., 2009, Niederle and Vesterlund, 2011, Sutter and Glätzle-Rützler, 2015, Almås et al., 2016, Iriberri and Rey-Biel, 2019).

and the Green type with a 50% chance. The reaction of types in MIX closely corresponds to this incentive. Women with the color green (W_G) have the highest and a significant rise in selection into tournament, followed by women with the color blue (W_B). There is a decrease in the competition entry of men of both colors (M_G and M_B) in MIX compared to the same types in CTR (test results reported in the figure note).

Stage 4 accounts for effects of preference for performing. In this stage, subjects submit their stage 1 performance to either tournament or piece-rate incentive without having to perform. Overall, competition entry in stage 4 is lower than in stage 3 in all treatments. The size of decrease is larger for women than for men in CTR and MIX, while it is larger for men than for women in GQ (see Figure 3.2). This evidence suggests that the presence or absence of performance does not necessarily result in variance in willingness to compete that is specific to gender.

Willingness to compete at the individual level

Next, we investigate determinants of willingness to compete at the individual level using parametric estimates. Table 3.2 shows the results of three probit regressions estimating the determinants of competition entry. All models estimate the effect of the treatments on the choice between tournament incentive and piece-rate incentive in stage 3. The predicted variable is a binary dummy, taking a value of 1 if a subject chooses the tournament incentive and 0 if a subject chooses the piece-rate incentive. We regress competition entry on treatment dummies (GQ and MIX) with CTR as the omitted group.

Model 1 regresses competition entry on treatment dummies with CTR as the base category and belief on own ranking in stage 3 as the control variable. In Model 2, we add risk attitude, two additional measures of overconfidence, and fairness perception of the policy in use as further controls. Model 3 estimates heterogeneous effects at type level, adding an interaction term between the dummy for being favored and the treatment dummies.[16]

The first model shows that willingness to compete is largely driven by belief on own rank. On a scale from 1 to 6 with 1 as the best rank, one unit decrease in belief on rank lowers the likelihood of entering the tournament by 10.4 percentage points. Model 2 additionally shows that risk attitude is a further drive for contest entry. On an 11-point Likert scale, one unit increase in willingness

[16] Table 3.5 provides detailed results and descriptions of the control variables. In this table, we also report the results of a version of Model 3, with an interaction term between risk and the dummy for the favored types.

to take risk increases the likelihood of competition entry by roughly 5 percentage points. Both gender quotas and mixed quotas have insignificant effects on competition entry in these models. When the indicator of being favored is included in Model 3, the result shows that mixed quotas significantly raise the probability to enter competition of the favored types by 34.0 percentage points. This effect under gender quotas is much lower, around 12.9 percentage points, and not significant.

Regarding control variables, we show that fairness perception and indirect measures of confidence matter for competition entry.[17] Results are robust if we use belief of rank in stage 2 instead of belief in stage 3, considering the argument that the decision to enter competition at the beginning of stage 3 is rather influenced by belief in the immediate previous stage 2.

Result 2a: A gender quota does not encourage women to enter competition while a mixed quota increases the competition entry of the targeted groups.

Beliefs, self-stigmatization, and willingness to compete

What leads to the seemingly surprising ineffectiveness of gender quotas? Our data provide evidence that gender quotas lower women's belief of relative performance and in turn reduce their willingness to compete. More importantly, this low level of confidence is not backed up by an actual difference in both absolute and relative performance.

On a scale from 1 to 6, with 1 as the best rank, the average belief in stage 3 of women in GQ is 3.10, higher than the average belief of women in CTR at 2.97 (Mann-Whitney U test, p= 0.5043), significantly higher than the average belief of the favored types in MIX at 2.61 (Mann-Whitney U test, p= 0.053), and marginally significantly higher than the average belief of women in MIX at 2.67 (Mann-Whitney U test, p= 0.105).

The difference in confidence does not correspond to differences in performance. Women in GQ solve, on average, 10.12 tasks correctly while in MIX

[17] For a more detailed analysis of fairness perception, refer to subsection 3.A.2. Gillen et al., 2019 replicates the experiment of Niederle and Vesterlund, 2007 and show that the gender gap in willingness to compete from this design is prone to measurement error. When several measures of risk attitude and confidence are added to the estimation, this gap becomes insignificant. We take into consideration their findings and elicit two measures of risk preference and three measures of confidence. Two measures of risk preference include a qualitative measure and an incentivized choice list. Three measures of confidence include belief of rank in the arithmetic task, belief of performance in the Cognitive Reflection Test (Frederick, 2005), and perception of mathematical ability. We refer readers to section 3.B and section 3.D for the detailed measures.

TABLE 3.2: Willingness to compete at the individual level

	(1)	(2)	(3)
GQ	-0.071	-0.007	-0.083
	(0.078)	(0.078)	(0.111)
MIX	0.043	0.100	-0.142
	(0.077)	(0.076)	(0.122)
GQ × Favored			0.129
			(0.146)
MIX × Favored			0.340**
			(0.148)
Favored			-0.084
			(0.110)
Belief of rank	-0.104***	-0.066***	-0.061***
	(0.020)	(0.022)	(0.022)
Risk measure		0.053***	0.054***
		(0.010)	(0.010)
Controls	No	Yes	Yes
N	240	240	240
Pseudo R^2	0.077	0.191	0.211

Notes: Average marginal effects of probit regressions estimating the likelihood of selecting the tournament incentive in stage 3. Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The binary dependent variable is willingness to compete in stage 3 (1 if choosing tournament incentive, 0 if choosing piece-rate incentive). *Favored* is a dummy variable (1 if favored type, 0 otherwise). *Belief of rank* is a categorical variable about beliefs of own rank (between 1 and 6, with 1 as the best rank) in stage 3; *Risk measure* is the answer to the general risk question elicited on an 11-point scale, higher numbers indicate a higher willingness to take risks. Table 3.5 provides detailed results and descriptions of the control variables.

favored subjects solve an average of 11.12 tasks correctly in stage 3 (Mann-Whitney U test, p= 0.358), and women solve an average of 11.22 tasks correctly (Mann-Whitney U test, p= 0.174) (see Figure 3.6).

In Table 3.3, we regress beliefs on treatment dummies. Model 1 controls for task performance in stage 3, and shows that gender quotas significantly increase belief (implying a worse perceived rank) by 0.324. When additional controls are taken into consideration in Model 2, the effect size increases to 0.405. Model 3 further disentangles this effect for the (non)favored types. Gender quotas significantly increase the beliefs of both men (by 0.700) and women (by 0.430). Mixed quotas increase the beliefs of the favored types by 0.104, and of the non-favored type by 0.143, yet both effects are not significant.

TABLE 3.3: The effect of quotas on beliefs

	(1)	(2)	(3)
GQ	0.324*	0.405***	0.700**
	(0.163)	(0.140)	(0.264)
MIX	-0.071	-0.045	0.143
	(0.153)	(0.132)	(0.270)
GQ × Favored			0.430*
			(0.240)
MIX × Favored			0.104
			(0.192)
Favored			0.270
			(0.287)
Stage 3 performance	-0.166***	-0.121***	-0.121***
	(0.018)	(0.020)	(0.021)
Constant	4.502***	4.649***	4.649***
	(0.221)	(0.286)	(0.286)
Controls	No	Yes	Yes
N	240	240	240
R^2	0.259	0.357	0.357

Notes: OLS estimation on effect of gender quotas and mixed quotas on belief on rank in stage 3. Clustered standard errors at the group level in parentheses. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. The dependent variable is a categorical variable about beliefs of own rank (between 1 and 6, with 1 as the best rank) in stage 3. *Favored* is a dummy variable (1 if favored type, 0 otherwise). Table 3.6 provides detailed results and descriptions of the control variables.

Overall, these results suggest that gender quotas lead to self-stigmatization. Women under gender quotas are less confident in their relative performance, thereby become less willing to enter competition. In contrast, mixed quotas do not seem to be prone to such consequences on the targeted groups.

Result 2b: A gender quota leads women to self-stigmatize their competence, thereby lowering their willingness to compete.

3.4.3 Post-competition cooperation

In this section, we first analyze perceived competence in team selection, including leader selection and colleague selection. Then, we look at the cost of misselection in terms of team performance.

Team selection

Figure 3.3 shows the average number of votes for leaders (left panel) and colleagues (right panel) by gender and treatment. We observe an overall tendency of favoring male leaders over female leaders in all treatments. In CTR, female leaders receive on average 1.55 votes while male leaders receive an average of 2.45 votes (Mann-Whitney U test, p= 0.099). This gap is *bigger and significant* in GQ while *smaller and insignificant* in MIX. On average, there are 1.31 votes for female leaders and 2.92 votes for male leaders in GQ (Mann-Whitney U test, p= 0.002); 1.75 votes for female leaders and 2.29 votes for male leaders in MIX (Mann-Whitney U test, p= 0.142). The difference in votes for male (female) leaders across treatments does not reach statistical significance.

In MIX, leaders from the Green (favored) group receive 0.6 votes on average, compared to 0.73 votes for leaders from the Blue (nonfavored) group (Mann-Whitney U test, p= 0.272), and 0.52 (0.67) votes for the Green group in CTR (GQ) (Mann-Whitney U tests, CTR vs. MIX: p= 0.796, GQ vs. MIX: p=0.192).

On the composition of votes, we observe an out-group favoritism regarding gender in CTR. Female leaders receive on average 1.8 times more votes from male than female group members, while this pattern reverses for male leaders who receive on average 1.5 times more votes from female compared to male group members. The change in the average number of votes for both male and female leaders between CTR and GQ is largely driven by changes in votes by male group members. In particular, female (male) leaders in CTR receive 1.3 (1.5) times more (less) votes from male group members compared to female (male) leaders in GQ. Cross-gender voting decreases, but does not completely disappear in GQ.

In contrast, we observe an in-group favoritism regarding gender in MIX. Female leaders in MIX receive more votes from women than men. On average, they receive 1.9 times more votes from female group members compared to female leaders in CTR. Male leaders in MIX receive 1.4 times more votes from male and 1.7 times fewer votes from female group members compared to male leaders in CTR.

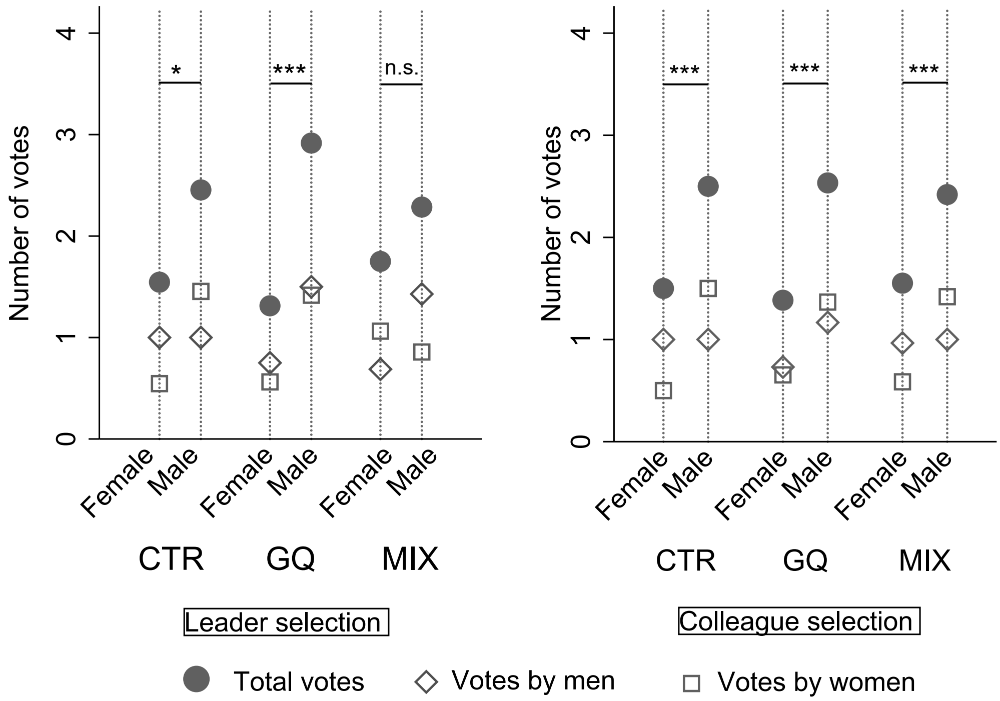

FIGURE 3.3: **Team selection**

Notes: the figure displays the average number of votes for leaders (left panel) and non-leaders (right panel) in stage 5. Female (Male) leaders receive 1.55 (2.45), 1.31 (2.91), 1.75 (2.29) votes in CTR, GQ and MIX, respectively (Mann-Whitney U test comparing votes for female versus male leaders, p= 0.099, p= 0.002, p= 0.142 for CTR, GQ and MIX respectively). Female (Male) colleagues receive 1.50 (2.50), 1.39 (2.53) 1.55 (2.42) votes in CTR, GQ and MIX, respectively (Mann-Whitney U test comparing votes for female versus male non-leaders, $p < 0.001$ for all treatments). The brackets and stars above each line show results of Mann-Whitney U test, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The absence of stars shows that a difference is not significant.

A similar tendency holds for colleague selection. Women are selected significantly less than men in all treatments, with a larger gap in GQ and a smaller gap in MIX compared to CTR (test results reported in the figure note). In terms of vote composition, an out-group favoritism regarding gender exists in all treatments. Women receive more votes from men than from other women, while men receive more votes from women than from other men.

To obtain a clearer picture, we take the gender composition of leaders into consideration. Table 3.4 describes this composition for each treatment. We focus on groups with one leader of each gender, which account for 63.64%, 85.71%, and 80.00% of the sample size in CTR, GQ, and MIX respectively. This is the most pointed case as each non-leader chooses between a male leader and a female leader. For these groups, the pattern observed in the whole sample becomes even more prominent in terms of size. Female leaders are voted for less frequently than male leaders in CTR. This gap is wider in GQ and narrower in MIX (see Figure 3.7 for specific pairwise comparisons).

TABLE 3.4: Number of observations and gender composition of leaders by treatment.

	CTR		Gender quotas		Mixed quotas		Total
Two male leaders	12	(18.18%)	0	(0%)	6	(6.67%)	18
Two female leaders	12	(18.18%)	12	(14.29%)	12	(13.33%)	36
One female and one male leader	42	(63.64%)	72	(85.71%)	72	(80.00%)	186
Total	66		84		90		240

Result 3: Without affirmative action, male leaders are perceived as more competent than female leaders. Gender quotas worsen the perceived competence gap between male and female leaders, while mixed quotas lessen it.

Result 3 points to the unintentional "stigma of incompetence" outcome of gender quotas, where women under this policy are perceived as less competent by others (Heilman et al., 1992). In stage 5, affirmative action is no longer in place and subjects work on a new unrelated task. Given this setup, result 3 suggests that the difference in the presumed competence of women is based on priors formed by previous exposure to different treatments. In addition, the general tendency of favoring male leaders in all treatments might be because the arithmetic task is perceived as a typical "stereotypical-male" task. In the next section, we look into grid task performance to test whether such beliefs are precise or erroneous.

Performance in team and the cost of misselection

Regarding performance in grid task, there are no significant differences in performance across treatments. Women perform better than men, yet the differences are not significant (see Figure 3.8 for specific pairwise comparisons). This result confirms the false perception of competence in team selection toward women observed in the previous section.

To study how team (mis)selection affects team performance. We calculate the deviation in percentage points between realized team performance and optimal team performance.[18] The realized performance is team performance created by

[18]For example, efficiency loss due to leader selection is calculated as (1 - (Performance of chosen leader/Performance of the best leader)) $\times$ 100%.

real team selection decisions, while the optimal team performance is calculated by assuming every subject selects their team optimally. In Figure 3.9, we show the inefficiency due to team selection overall, and disaggregate this measure into leader selection and colleague selection.

We observe no significant differences in overall inefficiency and in inefficiency due to colleague selection across treatments.[19] Inefficiency due to leader selection is highest in CTR (25.09%), lower in GQ (15.19%), and lowest in MIX (11.30%). For both men and women, the efficiency loss because of leader selection in MIX is significantly lower than in CTR (Mann-Whitney U test, p= 0.042, p= 0.043 for women and men, respectively). The differences between GQ and CTR are not significant (Mann-Whitney U test, p= 0.286, p= 0.104 for women and men, respectively). This evidence complements Result 3 on the reduced misperception of competence against female leaders in MIX.

Overall, gender quotas and mixed quotas do not harm cooperation measured as effort exerted in teamwork. However, efficiency loss emerges in all treatments if team misselection is considered. In line with Result 3, mixed quotas generate the least inefficiency caused by leader misselection.

3.5 Conclusion

In an increasingly diverse population and workforce, institutions have adapted their affirmative action policies to target multiple disadvantaged groups. In many cases, more than one affirmative action policy is implemented. Although decision-makers are aware of how each policy influences competition outcomes, applicants are usually uninformed thereof. For applicants, this feature entails an uncertainty about the actual favored group. We are the first to study how this feature affects outcomes of affirmative action policies in the form of quota rules in a laboratory experiment. We vary the rules of affirmative action, with one treatment favoring women with certainty and another treatment favoring either women or a member of one arbitrarily assigned group with a fifty-fifty probability. We compare both immediate and spillover outcomes of each rule with outcomes of a baseline treatment where affirmative action is not implemented.

[19]The overall inefficiency is highest in CTR (17.76%), followed by GQ (15.25%) and MIX (15.24%) (Mann-Whitney U test, GQ vs. CTR: p= 0.462, GQ vs. MIX: p= 0.797, and CTR vs. MIX: p= 0.400). Inefficiency due to colleague selection is highest in MIX (16.58%), then GQ (16.05%) and CTR (13.43%) (Mann-Whitney U test, GQ vs. CTR: p= 0.552, GQ vs. MIX: p= 0.730, and CTR vs. MIX: p= 0.323).

We found that independent of the policy in use, efficiency measured by task performance is not affected. This result is in line with previous literature (Balafoutas et al., 2016; Balafoutas and Sutter, 2012; Niederle et al., 2013), confirming that affirmative action is not harmful to efficiency. Interestingly, gender quotas fail to encourage women to enter the competition. We argue that the mechanism at work is self-stigmatization. Gender quotas lower women's confidence in their relative performance, hence lowering their willingness to compete. In contrast, mixed quotas raise the tournament entry of both favored groups.

Regarding indirect outcomes, our results are quite remarkable. We show that when affirmative action policy targets solely gender and is explicitly communicated as such, the policy seems to activate a stereotype threat about the competence of the favored group. Women are perceived as less competent in a subsequent teamwork. The uncertainty about the favored group reverses this effect. Putting our findings in perspective with the related literature, negative consequences of gender quotas observed in previous studies (Ip et al., 2020; Leibbrandt et al., 2017) might capture an upper bound of unintended effect because women are the only favored group in these studies, and their group identity is made salient.

The implementation of group identities in our designs is useful for analysis yet subject to certain limitations. First, the "color" characteristic might better resemble unobservable identities, such as socio-economic status or religion, but not other identities that are easy to make salient, such as age or disability. Second, we study a rather specific case when both identities are favored by affirmative action. Further research with more complex group identities, e.g, additionally including a group without any favored characteristics could provide interesting insights.

Overall, our results provide evidence that uncertainty over the targeted group and the salience of group identity do matter for both immediate and spillover outcomes of affirmative action policies. More generally, we illustrate how social identity is malleable under an institution, and in turn, affects the effectiveness of this institution. In organizational contexts, managers and policymakers should simultaneously target several disadvantaged groups and explicitly communicate affirmative action as such. At the same time, they might want to avoid strengthening the identity of selected candidates as being the favored or the unfavored.

3.A Additional Results

3.A.1 Performance under piece-rate and tournament incentives

Figure 3.4 displays performance under piece-rate incentives and tournament incentives by gender and treatment. In stage 1, women (men) solve on average 8.73 (9.91) tasks correctly in CTR compared to 8.93 (10.17) GQ and 9.37 (10.13) in MIX. The corresponding p-values of Mann-Whitney U tests comparing performance between men and women are p= 0.128, p= 0.253, p= 0.939 for performance in stage 1, p= 0.238, p= 0.105, p= 0.997 for performance in stage 2 for CTR, GQ and MIX, respectively.

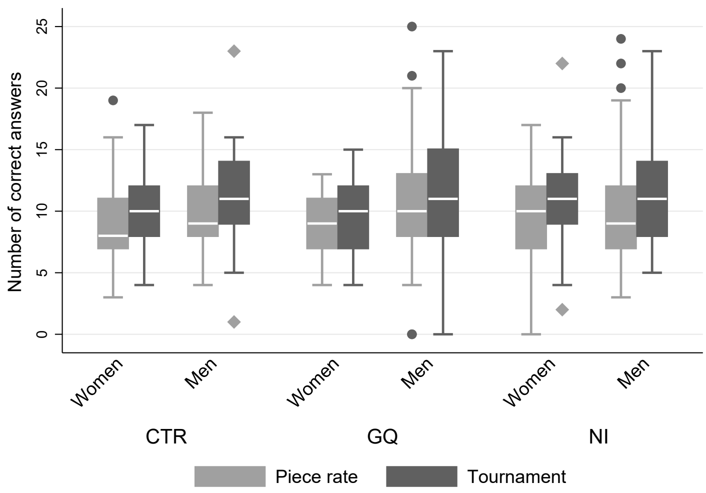

FIGURE 3.4: **Performance in stage 1 and stage 2 by treatment and gender**

Notes: Boxplots of performance in stage 1 (piece-rate) and state 2 (tournament) by gender and treatment. The upper (lower) hinges of the boxes show the 75th (25th) percentiles, the white lines inside the boxes show the median values, the upper (lower) adjacent lines show the maximum (minimum) and the dots show outliers.

3.A.2 Fairness perception

The left panel of Figure 3.5 shows the average perceived fairness for three policies under study by gender and treatment. In all treatments, women generally perceive gender quotas as fairer than men do (Mann-Whitney U test comparing average perceived fairness for gender quotas between men and women, $p < 0.001$, p= 0.083, $p < 0.001$ for CTR, GQ and MIX respectively). The gender gap in fairness perception for gender quotas is the largest in GQ. On one hand, women in GQ perceive gender quotas as fairer than women in other treatments, suggesting a self-serving bias (Mann-Whitney U test comparing the average perceived fairness for gender quotas among women across treatments, GQ vs. CTR:

p= 0.051, GQ vs. MIX: p= 0.469, and CTR vs. MIX: p= 0.187). On the other hand, men in GQ perceive this policy as less fair compared to men in other treatments (Mann-Whitney U test comparing the average perceived fairness for gender quotas among men across treatments, GQ vs. CTR: p= 0.895, GQ vs. MIX: p= 0.847, and CTR vs. MIX: p= 0.905). Intuitively, women as the targeted group of gender quotas in treatment GQ might anticipate that this policy is viewed unfavorably by men and consider their advantage as “unjustified", thus being reluctant to enter competition.

The right panel of Figure 3.5 shows the average perceived fairness for three policies under study for the (non)favored types in MIX and their reference types (same types) in CTR and GQ. The gap in fairness perception for mixed quotas between the favored and the nonfavored types is not the largest in MIX, but it is between the reference types in GQ, suggesting that mixed quotas are relatively “accepted" by the nonfavored type in MIX.

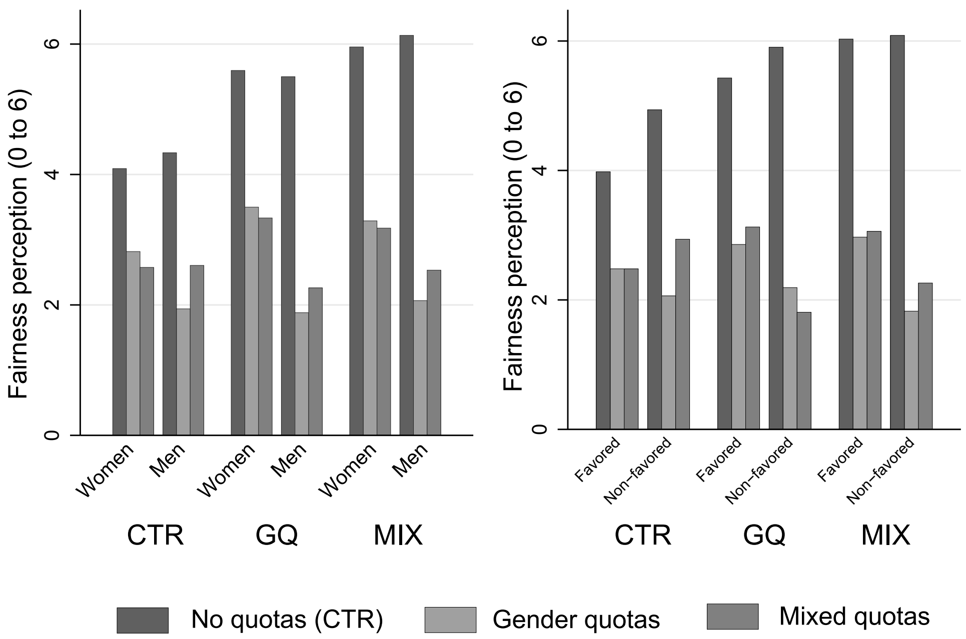

FIGURE 3.5: **Fairness perception**

Notes: The left panel of the figure displays the average fairness perceptions for different policies by gender and treatment. The right panel of the figure displays the average perceived fairness for three policies under study for the (non)favored types in MIX and their reference types in CTR and GQ. The favored types are Women_Blue, Women_Green, Men_Green; and the nonfavored type is Men_Blue. Higher numbers indicate that a policy is perceived as fairer.

3.A.3 Willingness to compete at the individual level

TABLE 3.5: Willingness to compete at the individual level - details

	(1)	(2)	(3)	(4)
GQ	-0.072	-0.007	-0.083	-0.085
	(0.079)	(0.078)	(0.111)	(0.112)
MIX	0.043	0.100	-0.142	-0.142
	(0.077)	(0.076)	(0.122)	(0.122)
GQ × Favored			0.129	0.130
			(0.146)	(0.146)
MIX × Favored			0.340**	0.342**
			(0.148)	(0.147)
Favored			-0.084	0.036
			(0.110)	(0.188)
Belief of rank	-0.104***	-0.066***	-0.061***	-0.063***
	(0.020)	(0.022)	(0.022)	(0.022)
Risk measure		0.053***	0.054***	0.066***
		(0.010)	(0.010)	(0.018)
Risk measure × Favored				-0.020
				(0.025)
Fairness perception		0.046***	0.037**	0.038**
		(0.017)	(0.018)	(0.018)
Belief of CRT score		0.098**	0.109**	0.104**
		(0.046)	(0.046)	(0.046)
Belief on math ability		0.021*	0.022**	0.022**
		(0.011)	(0.011)	(0.011)
N	240	240	240	240
Pseudo R^2	0.077	0.191	0.211	0.213

Notes: Average marginal effects of probit regressions estimating the likelihood of selecting tournament incentive in stage 3. Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The binary dependent variable is willingness to compete in stage 3 (1 if tournament, 0 if piece-rate). *Favored* is a dummy variable (1 if favored type, 0 otherwise). *Belief of rank* is a categorical variable about beliefs of own rank (between 1 and 6, with 1 as the best rank) in stage 3; *Risk attitude* is the answer to the general risk question elicited on an 11-point scale, higher numbers indicate a higher willingness to take risks; *Fairness perception* reflects fairness rating of own treatment, elicited on a 7-point scale on which higher numbers indicate higher perceived fairness; *Belief of CTR score* is belief of own rank in the Cognitive Reflection Test (Frederick, 2005), higher value indicates higher rank; *Belief on math ability* is the agreement on the statement "I am good at math", elicited on a 11-point scale, higher value indicates higher degree of agreement.

3.A.4 The effects of quotas on beliefs

Figure 3.6 displays the distributions of the difference between perceived ranks and actual ranks in stage 3 of all subjects in CTR, women in GQ, women in MIX, and the Green type in MIX. A negative (positive) difference implies that subjects believe they have better (worse) ranks than they actually have, hence being overconfident (underconfident).

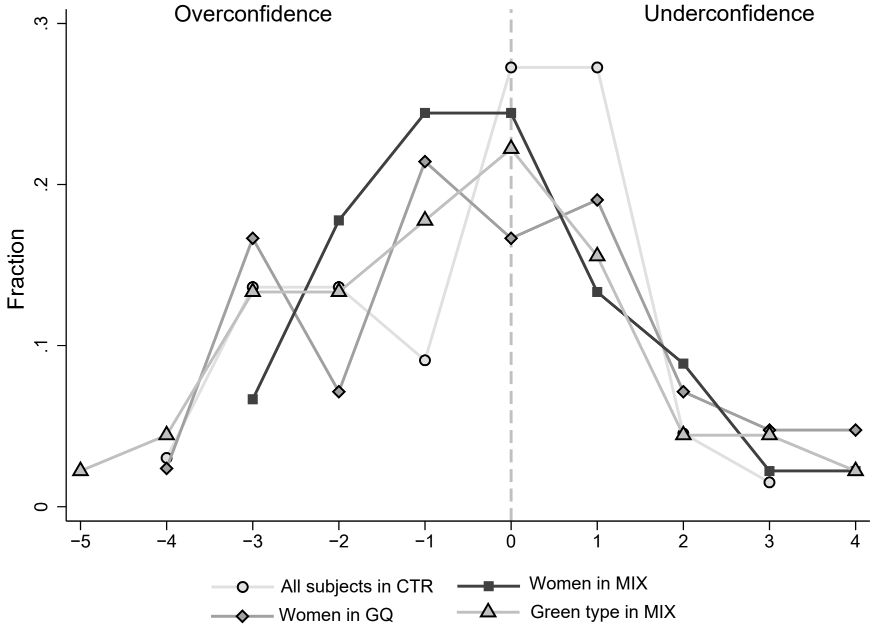

FIGURE 3.6: **Distribution of the difference between perceived rank and true rank**

Notes: The figure displays distributions of the difference between belief of own ranks and actual ranks in stage 3.

TABLE 3.6: The effect of quotas on beliefs - details

	(1)	(2)	(3)
GQ	0.324*	0.405***	0.700**
	(0.163)	(0.140)	(0.264)
MIX	-0.071	-0.045	0.143
	(0.153)	(0.132)	(0.270)
GQ × Favored			0.430*
			(0.240)
MIX × Favored			0.104
			(0.192)
Favored			0.270
			(0.287)
Stage 3 Performance	-0.166***	-0.121***	-0.121***
	(0.018)	(0.020)	(0.021)
Fairness perception		0.046	0.064
		(0.039)	(0.048)
Like task		-0.172***	-0.178***
		(0.044)	(0.042)
Belief on math ability		-0.064*	-0.062*
		(0.034)	(0.034)
Field of study		0.115**	0.112**
		(0.052)	(0.053)
Constant	4.502***	4.649***	4.649***
	(0.221)	(0.286)	(0.286)
N	240	240	240
R^2	0.259	0.357	0.357

Notes: OLS estimation on effect of affirmative action on belief of rank in stage 3. Clustered standard errors at the group level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable is a categorical variable about beliefs of own rank (between 1 and 6, with 1 as the best rank) in stage 3. *Favored* is a dummy variable (1 if favored type, 0 otherwise); *Fairness perception* reflects fairness rating of own treatment, elicited on a 7-point scale, higher numbers indicate higher fairness; *Like task* is the answer to the question "How much do you like the arithmetic task?" elicited on an 7-point scale, higher numbers indicate higher liking; *Belief in math ability* is the agreement on the statement "I am good at math", elicited on a 11-point scale, higher value indicates higher degree of agreement.

3.A.5 Team selection

Figure 3.7 displays the average number of votes for leaders (left panel) and non-leaders (right panel) in stage 5 for groups with one female leader and one male leader. Female (Male) leaders receive 1.29 (2.71), 1.08 (2.92), 1.67 (2.33) votes in CTR, GQ, and MIX, respectively (Mann-Whitney U test comparing votes for female versus male leaders, p= 0.023, p= 0.002, p= 0.050 for CTR, GQ, and MIX respectively). Female (Male) non-leaders receive 1.50 (2.50), 1.38 (2.63) 1.50 (2.50) votes in CTR, GQ, and MIX, respectively (Mann-Whitney U test comparing votes for female versus male non-leaders, p= 0.060, p= 0.002, p= 0.006 for CTR, GQ, and MIX respectively).

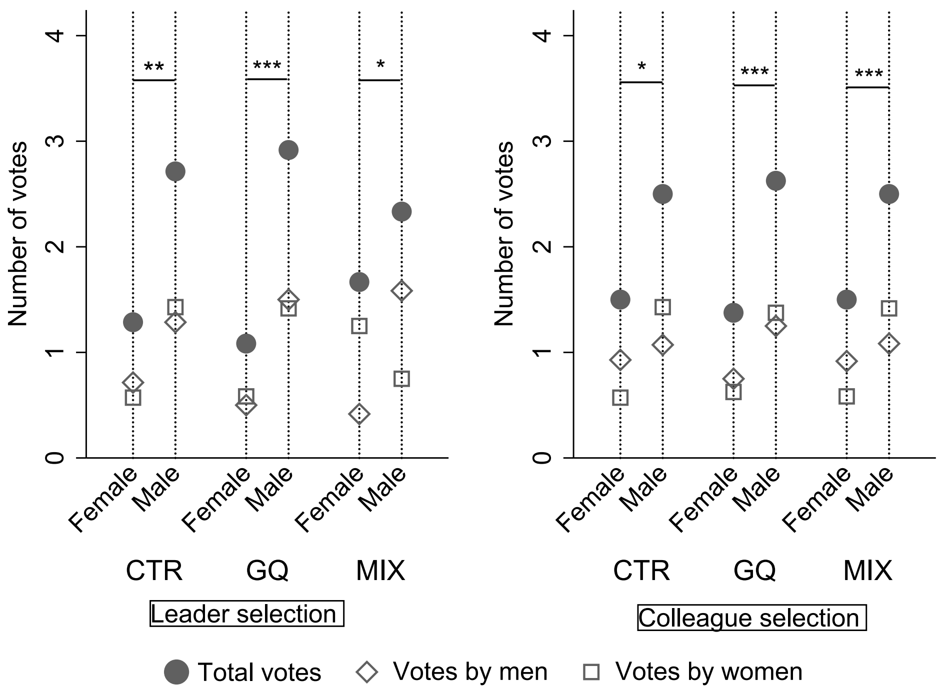

FIGURE 3.7: **Team selection in groups with one female leader and one male leader**

Notes: The figure displays the average number of votes for leaders (left panel) and non-leaders (right panel) in stage 5 in groups with one female leader and one male leader. The brackets and stars above each line show results of Mann-Whitney U test, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

3.A.6 Team performance

Figure 3.8 displays performance in grid task (stage 5) by gender and treatment. Subjects solve on average 6.77 grids correctly in CTR, 6.87 grids correctly in GQ and 7.32 grids correctly in MIX (Kruskal-Wallis test, p=0.359). Women (Men) solve 7.09 (6.46), 7.21 (6.52) and 7.76 (6.89) grids correctly in CTR, GQ, and MIX respectively (Mann-Whitney U test, p= 0.359, p= 0.240, p= 0.157 for CTR, GQ, and MIX respectively).

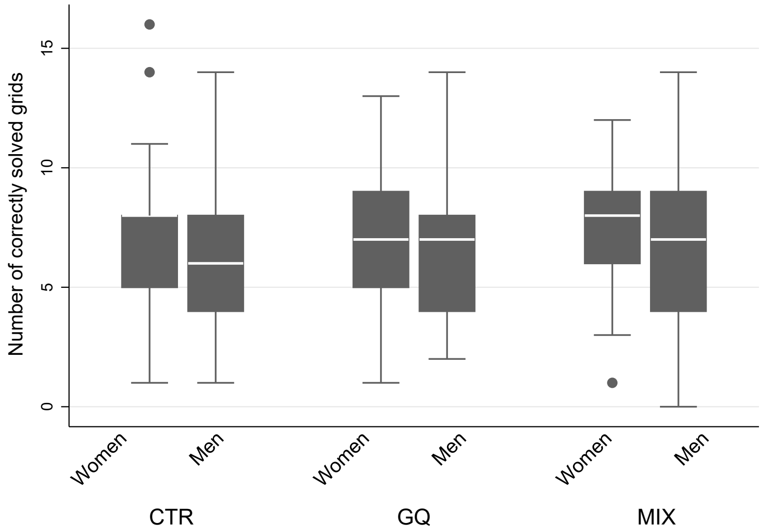

FIGURE 3.8: **Individual performance in grid task**

Notes: The figure displays the number of correctly solved grids in stage 5 of men and women in each treatment. The upper (lower) hinges of the boxes show the 75th (25th) percentiles, the white lines inside the boxes show the median values, the upper (lower) adjacent lines show the maximum (minimum) and the dots show outliers.

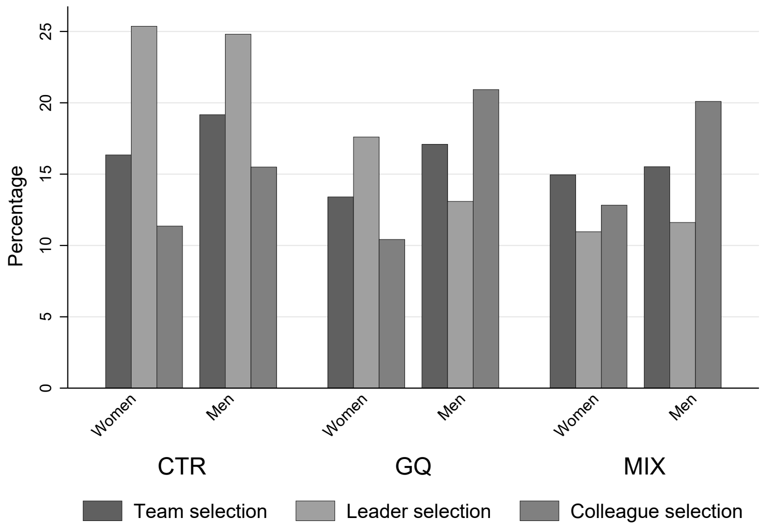

FIGURE 3.9: **Efficiency loss due to team selection**

Notes: The figure displays the deviation in percentage points between realized team performance and optimal team performance. For example, efficiency loss due to leader selection is calculated as (1 - (Performance of chosen leader/Performance of the best leader)) × 100%.

3.B Belief elicitation

(Translated from German, shown on screen at the end of stage 4)

In the following, we would like you to consider your performance in PART 1 (piece-rate), PART 2 (tournament), and PART 3 (choice), and guess your relative rank within your group of six. We also would like you to guess your relative rank among the three group members with the same gender/color as you.
Rank 1 is the group member with the highest performance, Rank 2 is the group member with the second highest performance, etc.
One of your guesses will be randomly chosen. If you are correct with this guess, you will also receive 1 EUR in addition to your payment from the other parts of the experiment. Thus, you should think carefully about each of the guesses.

- Which rank (from 1 to 6) within your group of six do you think you had in PART 1?
- Which rank (from 1 to 3) among the three group members with the same gender as you do you think you had in PART 1?
- Which rank (from 1 to 3) among the three group members with the same color as you do you think you had in PART 1?

Similar questions are asked about PART 2 and PART 3. The questions are programmed such that subjects can not give unreasonable answers. For example, a guessed rank of 1 in the whole group of six together with a guessed rank of 2 in group of same gender/color are not possible.

3.C Experimental instructions

(Translated from German)

General instructions (distributed on paper at the beginning of the experiment)

Welcome to today's experiment! Thank you for participating!

During the experiment, you and the other participants will be asked to make decisions. Your own decisions as well as the decisions of the other participants will determine your earnings, according to the rules that will be described in what follows.

The experiment will be conducted on the computer. You make your decisions on the screen. All your decisions and answers will remain confidential and anonymous.
The experiment consists of several parts. Additionally, you will answer a short questionnaire.
One of the parts will be selected randomly by the computer to determine your payment. Every part of the experiment is equally likely to be selected. It is therefore in your own interest to make your decisions in each part as if it was the only part.
Additionally you will receive a show-up fee of 4 EUR. This means that your total earnings will be the payment from the randomly chosen part plus the show-up fee of 4 EUR.
All other explanations will be given step-wise at the beginning of each part of the experiment. You will receive the instructions for each part in turn. You will have enough time to read the instructions carefully and to ask questions. Please do not hesitate to ask questions if something is unclear.
Please note that, as the last week, talking is not permitted. If you have questions, please do not ask them loudly but raise your hand. One of the experimenters will come to your seat to answer your question. If you do not comply with these rules you will be excluded from the experiment and you will not receive any payments.

General information regarding today's experiment

In today's experiment, your task is to add **as many five two-digit numbers as possible in a given amount of time.**
Each participant receive one color, either Green or Blue. The color will be assigned randomly to each participant. For **half of the participants**, the color will be **Green**. For **the other half**, it will be **Blue**.
We first start with a trial round. This round is for you to get used to the task. Therefore, your performance in this round is irrelevant for your final earnings. The trial round lasts 2 minutes. Afterwards, the first part of the experiment starts.
You will soon receive information about your color on the following screen.

PART 1 – Piece-rate (distributed on paper at the beginning of stage 1)

Your task in PART 1 is to add **as many five two-digit numbers as possible in five minutes.**

The use of a calculator or the similar is not allowed. You are allowed to use the provided scratch paper and pen if you like. After you have entered your answer, please click the “Confirm” button.
If PART 1 of the experiment is chosen for payment, you will receive the following payment:

$$\textbf{Payment = Number of correctly solved tasks} \times \textbf{0.50 EUR}$$

Your payment will not be reduced if you enter a wrong answer. We will refer to this payment as the **piece-rate payment** from now on.
You will be informed about your performance in this part at the end of the experiment.
After all questions regarding PART 1 are answered, your will start working on the task.

PART 2 – Tournament (distributed on paper at the beginning of stage 2)

As in PART 1, your task is to add as many five two-digit numbers as possible in five minutes. However, in this part your payment depends on your performance relative to the performance of other participants in your group.
Group allocation:
For the following parts of the experiment, you will be allocated to a **group with 6 members**. The groups were formed randomly and stay the same throughout the whole experiment. This means that you will form a group with the same participants for the rest of the experiment.
Each group consists of six **members**, and meets the following criteria:

- **Three** group members are women, **the other three** group members are men.
- **Three** group members are randomly assigned the color Green, **the other three** group members are randomly assigned the color Blue.

Rules of the tournament:
If PART 2 is chosen for payment, your payment depends on how high your performance is compared to the other five members of your group.
The **two** group members with **highest** performance are the two winners of the tournament.
(The content of the following part in gray differs across treatments. There is no further content for the control treatment (CTR).)

Gender quotas (GQ):
In addition, the following **special rule** is applied: **at least one winner must be a woman.**
If this is not automatically the case given the performance of the group members, then the female group member with the best performance among the three female group members will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**

Mixed quotas (MIX):
In addition, the following **special rule** is applied:

- With 50% probability, rule A is applied: **at least one winner must be a woman.**
- With 50% probability, rule B is applied: **at least one winner must be a group member with the color Green.**

Rule A: If this is not automatically the case given the performance of the group members, then the female group member with the best performance among the three female group members will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
Rule B: If this is not automatically the case given the performance of the group members, then the group member with the color Green and with the best performance among the three group members with the color Green will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**

The payment of the two winners is as follows:

Payment = Number of correctly solved tasks × 1.50 EUR

The **other four** members of your group get **no payment**.
If there is a tie between two group members, the winner will be determined randomly. We will refer to this payment as **tournament payment** from now on. At the end of the experiment, you will be informed about the outcome of the tournament.

In case you have any questions, please raise your hand.

PART 3 – Choice between piece-rate and tournament payment

(distributed on paper at the beginning of stage 3)
Similar to PART 1 and PART 2, your task is to add as many five two-digit numbers as possible in five minutes.
However, now you choose by yourself which payment scheme you prefer for your performance in PART 3. You can choose either the piece-rate payment (same rules as in PART 1) or the tournament payment (same rules as in PART 2).
If PART 3 is chosen for payment, your earnings will be determined as follows:

- If you choose the **piece-rate payment**, your payment is:

Payment = Number of correctly solved tasks $\times$ 0.50 EUR

- If you choose the **tournament payment, your earnings depend on the level of your performance in PART 3 compared to the performance of your five group members in PART 2 (tournament).** Reminder: PART 2 is the part you have just finished.

(The content of the following part in gray differs across treatments.)

Control treatment (CTR):
If your performance is higher than that of at least four other members of your group in PART 2, your payment is as follows:

Gender quotas (GQ): In addition, the following **special rule** is applied: **at least one winner must be a woman.**
If this is not automatically the case given the performance of the group members, then the female group member with the best performance among the three female group members will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
If your performance in PART 3 relative to the performance of your group members in PART 2 implies you are a winner, your payment is as follows:

Mixed quotas (MIX): In addition, the following **special rule** is applied:

- With 50% probability, rule A is applied: **at least one winner must be a woman.**
- With 50% probability, rule B is applied: **at least one winner must be a group member with the color Green.**

Rule A: If this is not automatically the case given the performance of the group members, then the female group member with the best performance among the three female group members will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
Rule B: If this is not automatically the case given the performance of the group members, then the group member with the color Green and with the best performance among the three group members with the color Green will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
If your performance in PART 3 relative to the performance of your group members in PART 2 implies you are a winner, your payment is as follows:

Payment = Number of correctly solved tasks $\times$ 1.50 EUR

That means it is three times as high as the piece-rate payment.
If your performance in PART 3 relative to the performance of the other group members in PART 2 implies that **you are not a winner**, you get **no payment**.
If there is a tie between two group members, the winner will be randomly determined.
The group composition is the same as in PART 2. If you choose the tournament payment, you will be informed about the outcome of the tournament at the end of the experiment.
On the next screen, you will decide whether you choose the piece-rate payment or the tournament payment for your performance in PART 3. Then the task will begin.
In case you have any questions, please raise your hand.

PART 4 – Choice between piece-rate and tournament payment for performance in PART 1 (distributed on paper at the beginning of stage 4)

In this part, you will not work on the task. Instead, you choose by yourself which payment scheme you prefer for your performance in PART 1. You can choose either the piece-rate payment (same rules as in PART 1) or the tournament payment (same rules as in PART 2) for your performance in PART 1.
If PART 4 is chosen for payment, your earnings will be determined as follows:

- If you choose the **piece-rate payment**, your payment is:

Payment = Number of correctly solved tasks in PART 1 $\times$ 0.50 EUR

- If you choose the **tournament payment, your earnings depend on the level of your performance in PART 1 compared to the performance of your five group members in PART 1.**

(The content of the following part in gray differs across treatments.)

Control treatment (CTR):
If your performance in PART 1 is higher than that of at least four other members of your group in PART 1, your payment is as follows:

Gender quotas (GQ): In addition, the following **special rule** is applied: **at least one winner must be a woman.**
If this is not automatically the case given the performance of the group members, then the female group member with the best performance among the three female group members will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
If your performance in PART 1 relative to the performance of your group members in PART 1 implies you are a winner, your payment is as follows:

Mixed quotas (MIX):
In addition, the following **special rule** is applied:

- With 50% probability, rule A is applied: **at least one winner must be a woman.**

- With 50% probability, rule B is applied: **at least one winner must be a group member with the color Green.**

Rule A: If this is not automatically the case given the performance of the group members, then the female group member with the best performance

among the three female group members will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
Rule B: If this is not automatically the case given the performance of the group members, then the group member with the color Green and with the best performance among the three group members with the color Green will **replace** the initial second-best winner. **In this case the group member with the second highest performance of all six group members of your group is no longer a winner.**
If your performance in PART 1 relative to the performance of your group members in PART 1 implies you are a winner, your payment is as follows:

Payment = Number of correctly solved tasks in PART 1 $\times$ 1.50 EUR

That means it is three times as high as the piece-rate payment.
If your performance in PART 1 relative to the performance of the other group members in PART 1 implies that **you are not a winner**, you get **no payment**.
If there is a tie between two group members, the winner will be randomly determined.
The group composition is the same as in PART 2. If you choose the tournament payment, you will be informed about the outcome of the tournament at the end of the experiment.
On the next screen, you will decide whether you choose the piece-rate payment or the tournament payment for your performance in PART 1.
In case you have any questions, please raise your hand.

PART 5 – Teamwork (distributed on paper at the beginning of stage 5)

In PART 5, you will work on a new task in which you have to solve as many counting tasks correctly as possible in five minutes, i.e. to correctly count the number of zeros ("0") in as many tables as possible. Each table consists of ten rows and ten columns, which contain either a zero ("0") or a one ("1"). Each table differs from the previous one. After you have entered your response, please click the "Confirm" button.
Leaders and colleagues:
Your group is divided into two leaders and four colleagues. The two leaders are the two winners in PART 2. The four colleagues are the four other group members who did not win in PART 2.
Team:

Each participant will form a team of three from their group of six. The team must consist of one leader and two colleagues. If you are a leader, you will choose two colleagues from your five group members for your team. If you are not a leader, you will choose one leader and one colleague from your five group members for your team.

You will not get any information about the identity of other group members and other group members will not get any information concerning your identity. The only information you will know about other group members before you make your decision is whether (s)he has won the tournament in PART 2 or not, and his (her) gender.

NOTE: Your team choice will not affect the team choice and the payment of other group members. The team choice of other group members will not affect your team choice and your payment. A leader/colleague can be chosen by more than one group member.

Payment:

Each member of your team of three will work on different tables of the counting task. Your payment in this part depends on how many correctly solved tables you and your teammates finish.

Precisely, your payment is determined as follows: you will receive 20 Euro-cent for each correctly solved table by each member of your team (including yourself). The other members of your team will also receive 20 Euro-cent for each correctly solved table that any team member (including yourself) has finished. This means each team member earns an equal amount of payment from the total correctly solved tables by the team (i.e. all three team members together).

In addition, the leaders (who won the tournament in PART 2) will each receive a bonus of 5 EUR. The colleagues will each receive a bonus of 2 EUR.

When PART 5 is chosen for payment, your payment is the sum of your bonus and your earnings (the sum of all correctly solved tables of your team × 20 Euro-Cent). At the end of the experiment, you will be informed about the overall performance of your team.

For example, if your team solve twenty tables correctly, the leader will receive the following payment:

$$\mathbf{20 \times 0.2 + 5 = 9\ EUR}$$

and the colleagues of your team will receive the following payment:

$$\mathbf{20 \times 0.2 + 2 = 6\ EUR}$$

In the following screen, you will make decision which group member are your teammates (leader and colleagues).

In case you have any questions, please raise your hand.

3.D Questionnaire

The questionnaire at the end of the experiment contains the following items (translated from German):

1. **Fairness perception:** In all treatments, we first elicit fairness perception of the competition rule in use, then of the other two competition rules, using the following scale:

   completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ *completely fair*

2. **Risk preference, general risk question:** same wording as in German Socio-Economic Panel questionnaire (SOEP, see, for example, Wagner et al., 2007) How do you evaluate yourself? Are you generally a risk-seeking person or do you try to avoid risks? The leftmost box means “not at all risk-seeking" and the rightmost “very risk-seeking". With the boxes in between, you can graduate your statement.

   not at all risk-seeking ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ *very risk-seeking*

3. **Risk preference, incentivized choice list:** Subjects make eleven, pairwise decisions between a lottery with a fifty-fifty chance of winning either 2 EUR or 7 EUR and a safe payment. The safe payment increases in 0.5 EUR increments, ranging from 2 EUR to 7 EUR.

4. **Cognitive reflection test:** see Frederick, 2005.

   After subjects finish the test, we elicit their belief about individual rank compared to the other participants in the experiment, with three possible answers: 0-33.3%, 33.4-66.6%, and 66.7-100% (implying they answer one/two/three out of three questions in the test correctly, respectively.)

5. **Social preference** (survey question, Falk et al., 2018)
   Question 1: Imagine the following situation: Today you unexpectedly received 1000 EUR. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed).
   Question 2: Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize that you lost your way. You ask a stranger for directions. The stranger offers to take

you to your destination. Helping you costs the stranger about 20 EUR in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 EUR, the most expensive one costs 30 EUR. Do you give one of the presents to the stranger as a "thank you" gift?
Which present do you give to the stranger?
1. No, would not give present
2. The present worth 5 EUR
3. The present worth 10 EUR
4. The present worth 15 EUR
5. The present worth 20 EUR
6. The present worth 25 EUR
7. The present worth 30 EUR

6. **Socio-demographics:** age, final grade point average at academic high school, last math grade at academic high school, field of study, monthly disposable amount of money, political orientation, number of experiments already participated in the same lab.

7. **Positive and negative reciprocity, trust, and belief on own mathematical ability** (survey question, Falk et al., 2018)

Chapter 4

Consumers with limited attention in the credence goods market

Co-authored with Vasilisa Petrischeva, Alexander Rasch, and Christian Waibel

4.1 Introduction

Credence goods markets are characterized by the distinct feature of informational asymmetries where sellers have an informational advantage over customers. Sellers are experts and know which type of service customers need, whereas customers do not. Experts may exploit this advantage by providing more or fewer services than necessary.[1] Many important markets such as repair service and healthcare are considered credence goods markets.

One of the key theoretical predictions is that experts should have no incentives to provide an inappropriate amount of service whenever customers can verify the type of service (Dulleck and Kerschbamer, 2006). In equilibrium, experts post equal markups for the different types of services. By posting equal markups, experts credibly signal performing the type of service that the customer needs. As customers anticipate that experts provide necessary services under equal mark-ups, customers' willingness to pay for a service is maximal. A monopolistic experts set these equal markup prices such that they can fully extract the rent. In a competitive credence goods market, the price covers experts' marginal costs of providing a service.

In real markets, however, these predictions appear to contradict observations. The FBI estimates that up to 10% of the 3.3 trillion US dollars of yearly health expenditures in the United States are due to fraud (Federal Bureau of Investigation, 2011).[2] Gottschalk et al., 2020 show that 28% of dentists' treatment recommendations involve overtreatment recommendations. In car repair services, Taylor, 1995, Schneider, 2012, and Rasch and Waibel, 2018 report fraud performed by garages. Kerschbamer et al., 2016 document fraud in computer repair services. And Balafoutas et al., 2013 as well as Balafoutas et al., 2015 identify fraud in the market for taxi rides.

In this paper, we aim to study whether such discrepancies between the theoretical results and real-life observations can be explained by consumers' limited attention. Our focus on limited attention is motivated by the fact that in many credence goods markets, there is a call to make experts' earnings more transparent. An example of a sector in which more transparency is demanded is the market for health services. In Germany, for instance, many health services are paid for by the patients' insurance companies. The payments are organized bilaterally between the insurance company and the physician without any patient

[1]See, e.g., Dulleck and Kerschbamer, 2006 for a theoretical framework, and Kerschbamer and Sutter, 2017 for an overview of experimental evidence.

[2]For an overview of the phenomenon of so-called physician-induced demand (PID), see McGuire, 2000

involvement. In order to increase transparency for such services, since 2012 patients have had the right to ask for a patient receipt. This receipt must report the treatments performed and the (expected) costs.[3]

Increased transparency can also be advocated by the providers themselves. For example, for their car repair services, car maker Opel introduced a new app-based information service called "MyDigitalService". When car owners have their cars inspected or repaired, they can now more easily follow the different steps in the process, and are provided with information regarding additional costs when unanticipated services become necessary.[4]

Employing a simple model, we predict that customers' limited attention increases the level of insufficient service and raises the markup difference between the major and the minor services. Customers are also more willing to pay for an offer that triggers insufficient service if they pay limited attention. In our model, customers suffer from either a minor or a major problem. The major service solves both problems but is more costly to the expert than the minor service, whereas the minor (and less costly) service can only solve the minor problem. Service costs are common knowledge among experts and customers. By posting an equal markup price vector, an expert can credibly signal that she has no incentive to over- or undertreat. Experts and customers are randomly rematched in our lab experiment and hence do not suffer from reputational concerns. Once an expert has posted her price vector, the customer observes the price vector.

We vary whether a customer observes – in addition to the expert's price vector – the expert's cost vector. A customer then decides whether or not he wants to interact upon the posted prices. The expert observes which type of problem her customer has and decides whether to provide the either the minor or the major service. The expert charges for the provided service.

We test the predictions in a laboratory experiment. In the treatment Salience, customers observe the prices and experts' costs are made salient before deciding on interaction, while in the NoSalience treatment customers only observe prices. We find that experts' price vectors are significantly closer to the equal markup when costs are made salient than when they are not. Customers' interaction probability decreases by around 20 percentage points over time, and does not significantly vary across treatments. Controlling for subjects' covariates, experts undertreat customers significantly more often under NoSalience than under Salience.

[3]See, e.g., Federal Ministry of Health, 2020

[4]See, e.g., automotorundsport.de, 2020.

Salience decreases total welfare calculated as accumulated profits. Due to the strong decrease in interaction over time, the customer surplus is smaller under Salience than under NoSalience. Experts benefit from limited attention because they can extract the additional surplus generated by more sufficient treatments. When welfare is calculated as accumulated profit minus the outside option, and random differences in the customers' type of problem (minor or major) are considered, welfare is improved under Salience.

Our study is directly related to the literature on credence goods. Closest to our paper is the article by Dulleck et al., 2011, which employs a large-scale laboratory experiment challenging the seminal model by Dulleck and Kerschbamer, 2006. In particular, the authors study the impact of institutions, such as verifiability or liability, on outcomes in credence goods markets and show that liability is an effective tool for improving outcomes in credence goods markets. However, the authors find no evidence that verifiability fosters market results.

There are multiple explanations for why verifiability seems to be less effective for improving market outcomes. Previous work has explained the differences between the prediction of no overtreatment if services are verifiable and the observation in real markets mostly based on experts' characteristics. Emons, 1997, 2001 argues that experts' utilization of capacities drives overtreatment. If demand is low, experts may have the incentive to provide excess services to fill capacities. Gottschalk et al., 2020 provides evidence from a field experiment. Dentists with a low utilization are correlated with a higher probability of receiving an overtreatment recommendation. Hilger, 2016 develops a model that accounts for experts' heterogeneity with respect to experts' costs of service provision. If costs are unobservable, experts cannot credibly signal equal markups. Hilger, 2016 assumes experts are liable for their services. Hence, experts may have an incentive to overtreat.

To our knowledge, the only paper that is based on customers' characteristics is Kerschbamer et al., 2017, which suggest that customers' preferences may drive the deviations observed in Dulleck et al., 2011. More precisely, the authors argue that a heterogeneity in social preferences may explain the observed behavior. They show theoretically that equal-price equilibria are robust to pro-social preferences but not to anti-social preferences. Our study extends this strand of literature by adding the perspective of consumers' limited attention.

Our paper also contributes to the literature on behavioral industrial organization, which investigates market outcomes when consumers have behavioral biases.[5] The closest strand of literature to our setup are studies on add-on pricing (Armstrong and Vickers, 2012; Gabaix and Laibson, 2006; Grubb, 2015b; Heidhues and Kőszegi, 2017), in which consumers do not pay attention to the additional price of a two-part tariff. Our study contributes by investigating limited attention on a different factor, namely sellers' costs. In particular, customers are fully attentive to the prices of two treatments offered by the sellers, but not to the cost of each treatment. Costs do not directly show up in the customers' payoff function, yet they influence the treatment offered by sellers. The chosen treatment then determine whether consumers receive a proper treatment, which then affects their payoffs. Compared to the literature on add-on pricing, our study allows to explore a rather indirect setup that is highly relevant in the context of credence goods markets.

The remainder of the paper is as follows. Section 4.2 provides the theoretical framework for the credence goods market. Section 4.3 lays out our experimental design and shows our hypotheses. Section 4.4 displays and discusses our results before we conclude in section 4.5.

4.2 Theoretical framework

4.2.1 Market

We model a market with verifiability and without liability (Dulleck and Kerschbamer, 2006).[6] Consider a market with an expert and a customer. A customer (she) has either a major or minor problem. The customer knows that she has a problem, but does not know whether it is major or minor. However, the customer knows that she has the major problem with an ex-ante probability h and the minor problem with an ex-ante probability $(1-h)$. These probabilities are common knowledge to both the expert and the customer.

The expert (he) can identify the problem at no cost. He can choose to provide either a major or a minor treatment. The cost of the major treatment is $\bar{c}$ and the cost of the minor treatment is $\underline{c}$, with $\underline{c} < \bar{c}$. The major treatment heals both problems, whereas the minor treatment only heals the minor problem. The customer has a valuation $v > 0$ of getting treated sufficiently. The expert is not liable – that is, he can treat a customer who has the major problem with a minor treatment. The prices for the minor and major treatment are denoted

[5] See, e.g., Grubb, 2015a and Heidhues and Kőszegi, 2018 for an overview.

[6] Verifiability implies that an expert cannot charge for the major treatment if he has provided the minor treatment. Liability implies that an expert cannot provide the minor treatment if the major treatment is needed (Dulleck and Kerschbamer, 2006).

as $\underline{p}$ and $\bar{p}$, respectively, with $\underline{p} < \bar{p}$. Due to the verifiability of a treatment, the expert has to charge $\underline{p}$ if he provides the minor treatment and $\bar{p}$ if he provides the major treatment (no overcharging). The customer does not know the necessary treatment, but knows whether she has been undertreated. The expert posts take-it-or-leave-it prices.

The game is characterized as follows:

1. The expert posts a price menu $(\bar{p}, \underline{p})$ for the major and minor treatment, respectively.

2. The customer chooses whether to interact where the presentation of information differs across treatments:

   (a) *NoSalience* treatment: The customer observes the price menu posted by the expert.

   (b) *Salience* treatment: The customer observes the price menu posted by the expert and the expert's (potential) profit for each price.[7]

   If the customer chooses not to interact, the game ends, and the expert and the customer both get the outside option.

3. If the customer chooses to interact, nature draws the type of problem that the customer has.[8]

4. The expert observes the problem type of the customer. The expert then provides either a major or a minor treatment, and charges a price according to his treatment recommendation ($\bar{p}$ or $\underline{p}$).

5. The expert observes his payoff, and the customer observes her payoff.

If there is interaction, the expert's payoff (profit) is determined by the price p ($p \in \{\underline{p}, \bar{p}\}$) minus the cost c ($c \in \{\underline{c}, \bar{c}\}$) of the treatment applied, i.e., $\pi_e = p - c$. If there is no interaction, the payoff amounts to u.

If the customer chooses to interact, and is not undertreated, she derives her valuation of v. If she decides to interact, and is undertreated, she derives a valuation of zero. In either case, the customer must pay the price p of the treatment

[7]Note that even when a customer cannot directly observe the expert's profit, she can calculate the profit, because the costs of both treatments are common knowledge.

[8]As Dulleck and Kerschbamer, 2006 point out, it does not make a (game-theoretic) difference whether nature determines the severity of the problem after the customer has consulted an expert (but before the expert has performed the diagnosis) or at the very beginning.

she receives. Hence, for each period, her payoff is $\pi_c = v - p$ if she is not undertreated, and $\pi_c = -p$ if she is undertreated. If the customer decides not to interact, she receives a payoff of u.

The game and the payoffs are illustrated in Figure 4.1.

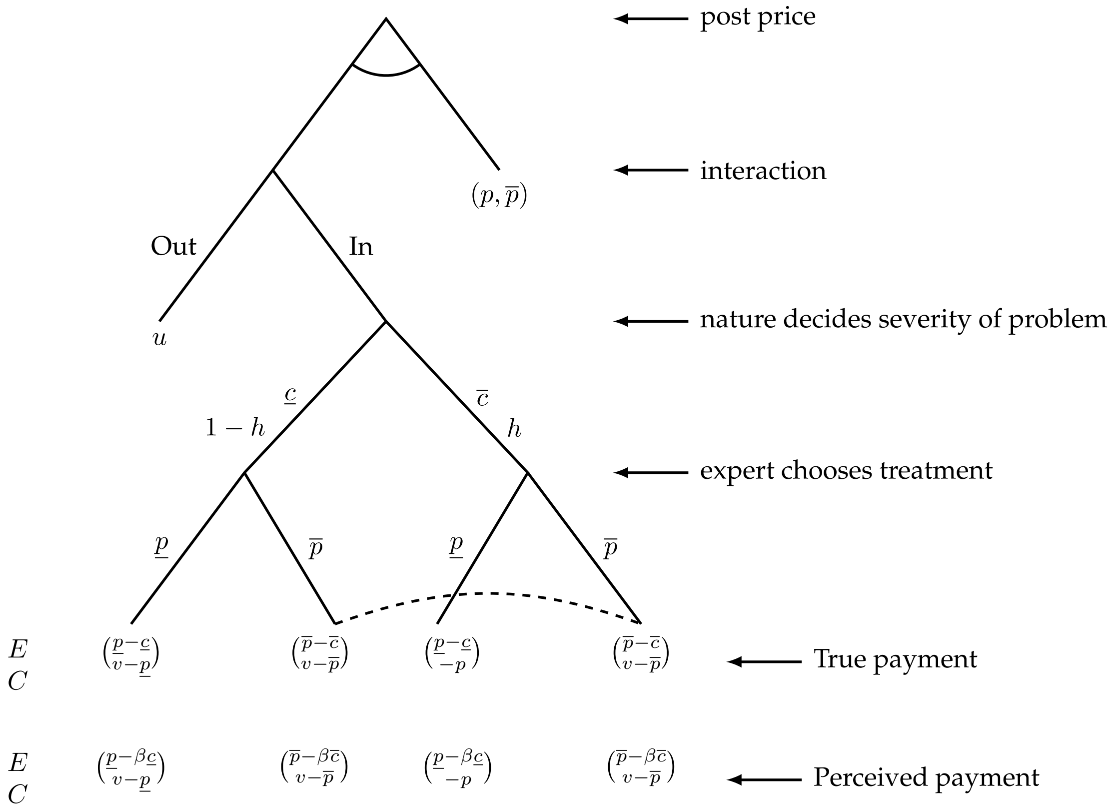

FIGURE 4.1: **Game tree: Market with limited-attention customers**

In the following, we will sometimes refer to an undertreatment price vector which is defined as a price vector with a negative markup difference. Because the minor treatment has a higher markup, it is more profitable for an expert to provide a minor treatment regardless of the severity of the customer's problem.

4.2.2 Customers with limited attention

We assume that the customers have limited attention. When deciding on interaction, there are three related features of this decision, namely prices $(\underline{p}, \bar{p})$, valuation v, and the likelihood of being undertreated. The likelihood of being undertreated is directly determined by whether the customer has a major or a minor problem, and the action that is chosen by the expert. As the expert is a profit-maximizing agent, we assume that the expert always chooses the action that gives him the higher profit.

We assume that prices and the valuation are salient features, whereas the probability of being undertreated is a hidden feature. This assumption is backed up by three observations. First, several laboratory experiments on credence good markets (see, e.g., Dulleck et al., 2011) have a design feature that shows only the prices to customers when they decide on interaction. Second, valuation and prices immediately show up in the customer's payoff function. Third, although the expert's profit function and costs are common knowledge, they are communicated to the customer once at the beginning of the experiment. Thus, it is reasonably more difficult for customers to recall this information in each and every period. When seeing the information of the expert's profit in Decision 3, the customer takes the hidden feature into consideration when deciding on interaction.

Because the expert's profit is equal to price minus cost, we consider expert's profit of each treatment as the direct proxies for the hidden feature of the likelihood of being undertreated. In the experiment, we manipulate the salience of this hidden feature by (not) showing expert's profits at the interaction stage, hence (not) indicating costs.

Finally, we assume that the expert is aware that the customers have limited attention, but the customer is not aware that the expert knows thereof.

The degree of limited attention is captured by parameter β ($\beta \in [0, 1]$) If $\beta \to 1$, all features are identically salient. If $\beta \to 0$, the customer takes only the salient features into consideration, and completely neglects the hidden feature.

If the customer decides to interact, the expert's profit is $\pi = p - c$, while profit as perceived by a customer with limited attention equals $\pi = p - \beta c$.

Lemma 1. *When the customer has limited attention to costs, the equal-markup tariff $(\bar{p}, \underline{p})$ is perceived by customers as a tariff, such that the markup for the major treatment exceeds that for the minor treatment.*

Proof. See section 4.A □

Proposition 1. *When the customer has limited attention to costs, conditional on interaction,*

(i) the expert always posts tariffs with the markup of the minor treatment exceeds that for the major treatment, and

(ii) the expert always provides the minor treatment.

Proof. See section 4.A □

4.3 Experiment

4.3.1 Experimental design

Our *NoSalience* treatment replicates the results from the baseline treatment with verifiability in Dulleck et al. (2011). We introduce salience of experts' costs and refer to the latter as the *Salience* treatment.

Subjects are assigned to be either an expert (called Player A in the experiment) or a customer (called Player B in the experiment). Each market consists of eight subjects, with four experts and four customers. In each period, one expert interacts with one customer. The assignment to market and role is random and does not change during the experiment. The stage game is repeated for 16 periods. Experts and customers are re-matched within their market at the beginning of each period. At the end of each period, subjects are informed about their profit for the current period, as well as their own accumulated profit. Subjects never learn about the profits of others.

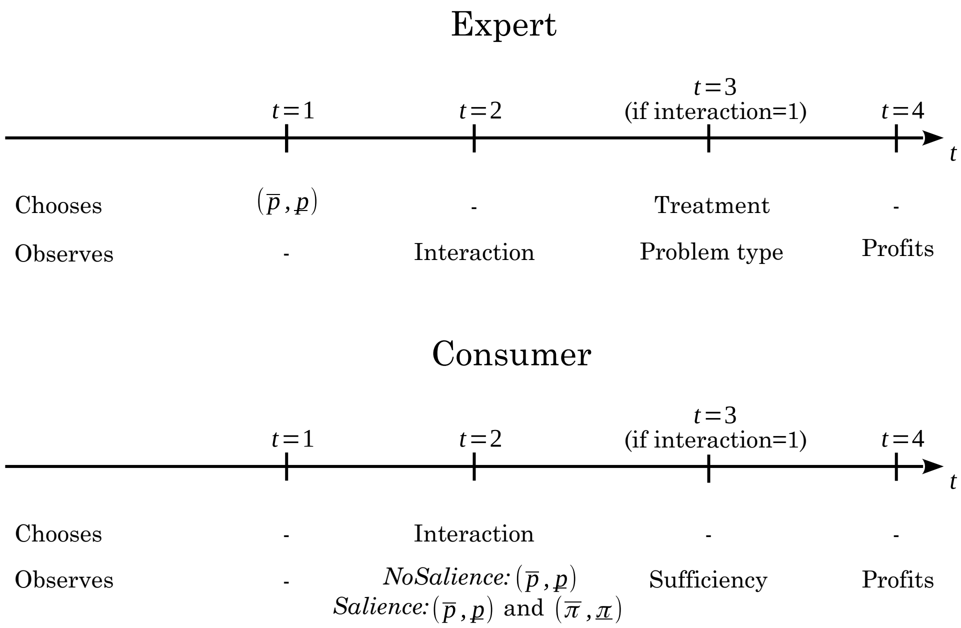

FIGURE 4.2: Timing of decisions

The timing is displayed in Figure 4.2. In each period, the expert chooses prices $p_i \in [1, 11] \in N$ for each of the two treatments. The customer then chooses whether or not to interact. If a customer decides to not interact, the period ends and she and her matched expert both get u = 1.6 ECU (outside option).[9] If a

[9] ECU refers to "Experimental Currency Unit", 1 ECU is equal to 25 Euro-cents.

customer decides to interact, the expert provides either the minor treatment $\underline{c}$ (called Action 1 in the experiment) at costs of 2 ECU or the major treatment at costs of 6 ECU $\bar{c}$ (called Action 2 in the experiment). The customer derives a utility v = 10 ECU if she is sufficiently treated and 0 otherwise. The probability of a customer having a major problem is $h = 0.5$.

After the experiment, we ask for individuals' beliefs conditional on the subjects' role of a buyer or a seller. We further use incentivized choice lists to elicit individual risk preference and loss aversion. We complement the incentivized decisions with a validated question on general risk preference. Selected questionnaire items from the preference survey module of Falk et al., 2018 serve as a measure of social preferences. We complete the post-experimental part by recording individuals' reasoning for their decisions in the experiment and their socio-demographics.[10]

Table 4.1 provides an overview of subjects' covariates. The left column shows the averages across all participants, the two middle columns show the same value per treatment, and the significance levels of the differences (two-sided Mann-Whitney U tests) are in the rightmost column. Based on the risk choice list, our subjects are, on average, slightly risk averse. Responses to the risk question is consistent with the incentivized choice list. In both measures, an average value that is smaller than 5 implies risk aversion. Subjects are loss averse with virtually no variation across treatments.

TABLE 4.1: Overview of the sample

	All	$NoSalience$	$Salience$	Difference
Risk preference (choice list)	4.22	4.23	4.22	p=0.953
Risk preference (question)	3.78	3.60	3.90	p=0.370
Loss aversion (choice list)	4.31	4.33	4.29	p=1.000
Social preference (question 1)	118.64	140.67	103.96	p=0.409
Social preference (question 2)	4.14	4.08	4.18	p=0.679
Gender	0.53	0.63	0.47	p=0.038
Age	24.72	25.15	24.43	p=0.345
N	120	48	72	

We account for the role of social preference (Kerschbamer and Sutter, 2017) by classifying individuals into two categories based on the second measure of

[10] We refer readers to section 4.D for details on each measure.

social preference.[11] We then define a market as pro-social based on the number of pro-social experts in this market (from 0 to 4) and treat this measure as a continuous control variable in all following regressions. We take into account only the number of pro-social experts because experts' decisions (choices of prices and treatment) might reflect his social preference. We consider this to be not the case for customers who can only accept or reject experts' offers.

4.3.2 Procedures

We conducted our experiment at the DICE Lab of the University of Düsseldorf in June 2019, using zTree (Fischbacher, 2007). Subjects were recruited via ORSEE (Greiner, 2004) and were mostly enrolled as students at the University of Düsseldorf. Upon arriving at the lab, each subject was randomly assigned to a cubicle and provided with instructions. Subjects were given enough time to read the instruction, and were allowed to ask the experimenters questions privately. The sessions started after all questions had been addressed.

In total, 120 subjects participated in six sessions of the experiment. Each session lasted for about 90 minutes. On average, subjects earned 18.34 EUR. In total, 48 subjects participated in *NoSalience* and 72 subjects participated in *Salience*.

4.3.3 Hypotheses

Based on our theoretical model and the experimental parameterization, we now form our hypotheses for customers' and experts' behaviors.

Hypothesis 1. *Customers are more likely to interact given an undertreatment price vector under NoSalience than under Salience.*

Hypothesis 2. *Experts are more likely to post an undertreatment tariff under NoSalience than under Salience.*

Hypothesis 3. *Experts' markup differences are larger under NoSalience than under Salience.*

Hypothesis 4. *Experts are more likely to undertreat customers under NoSalience than under Salience.*

[11]See Question 2 of social preference measures in section 4.D. We consider one subject as pro-social if she is willing to give at least the amount the stranger spent to help her (20 EUR or more), and as selfish otherwise (15 EUR and less).

4.4 Experimental Results

In this section, we first present the direct effect of salience, i.e., how it affects customers' decision on interaction. We then investigate the indirect effects, i.e., how salience influences experts' decisions by analyzing the price vectors that experts post and the treatment composition they provide given their posted prices. Finally, we discuss how salience influences customers' and experts' welfare. Throughout, we look in depth at how each outcome varies for different types of individuals and markets according to the social preference classification.

Table 4.2 provides a first overview on the aggregate level.

TABLE 4.2: Descriptive statistics

	Salience		*NoSalience*	
	Mean	Std. Dev.	Mean	Std. Dev.
Interaction	0.52	0.50	0.48	0.50
Major price $\overline{p}$	7.93	1.35	7.77	1.57
Minor price $\underline{p}$	5.39	1.73	5.68	1.67
Markup difference	-1.47	1.94	-1.91	1.69
Sufficient treatment	0.38	0.49	0.37	0.48
N	1152	1152	768	768

In the individual-level data analysis, we control for the price level and previous period market characteristics. We further account for individual experts' characteristics which include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs (see section 4.D).

4.4.1 Interaction

Interaction is directly affected by our treatment variation. Upon observing experts' price vectors (plus salience of costs in treatment *Salience*), customers decide whether or not to interact. The trade-off they face is whether to opt for a certain outside option of 1.6 ECU or to interact and face a risk of being mistreated. In treatment *Salience*, customers are supposed to be more attentive to the risk of being mistreated as experts' costs become more salient to them.

Figure 4.3 displays the interaction given different types of posted price vectors across treatments. The upper-left graph show that, overall, interaction in

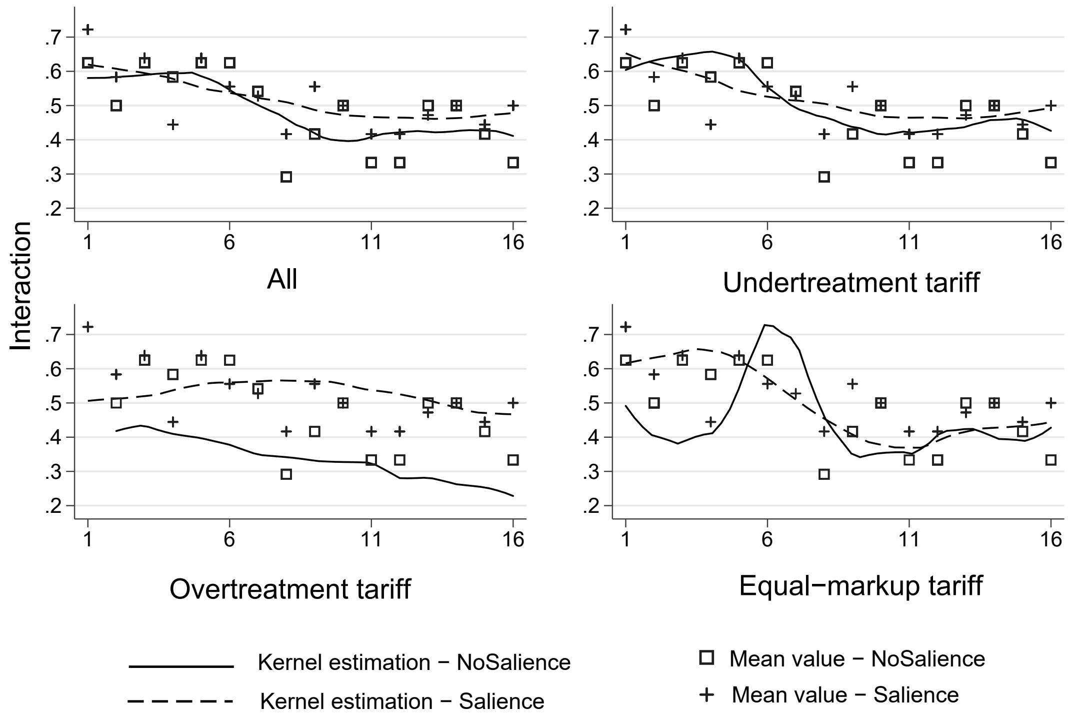

FIGURE 4.3: **Interaction over time by treatments and price vectors**

Note: The figure displays kernel density estimations and mean values of interaction over 16 periods by treatments. The fitted values are estimated using the Epanechniov kernel with the optimal bandwidth. The upper-left graph displays the interaction for all price vectors. The upper-right graph displays the interaction given an undertreatment price vector. The lower-left graph displays the interaction given an overtreatment price vector. The lower-right graph displays the interaction given an equal-markup price vector.

Salience is slightly higher than in *NoSalience*, and there is a declining tendency over time in both treatments.[12]

Overall, we do not find strong supportive evidence for Hypothesis 1. The estimations in the upper-right graph of Figure 4.3 shows that customers in *NoSalience* interact more given an undertreatment price vector at the beginning. From around period 6 onward, the interaction probability under *NoSalience* becomes slightly lower than under *Salience*. On the same note, although we might expect interaction given an equal-markup tariff is higher in *Salience* compared to *NoSalience*, this is not the case for all periods.[13] Interaction given equal-markup price vectors (the lower-right graph) in *Salience* is higher than in *NoSalience* only

[12] The average interaction rates of all periods are 48.44% and 52.08% in *NoSalience* and *Salience*, respectively (two-sided Mann-Whitney U test, p= 0.269). In the first period, 62.5% of customers in *NoSalience* and 72.2% of customers in *Salience* choose to interact, whereas in the last period, only 33.33% of customers in *NoSalience* and 50% of customers in *Salience* choose to do so.

[13] An equal-markup tariff is perceived as an overtreatment tariff by a customer with limited attention (see Table 4.9). Assuming customers are more attentive in *Salience* than in *NoSalience*, they are, on average, more likely to recognize an equal-markup tariff and hence, theoretically more likely to choose interaction compared to customers in *NoSalience*.

up until period 4, then decreases steadily and becomes less than *NoSalience* between periods 5 and 8, before slightly rising and becomes mostly higher than *NoSalience* again between periods 9 and 16.

On the other hand, the lower-left graph shows that customers in *Salience* consistently consider an overtreatment price vector as more attractive than customers in *NoSalience*. The interaction probability given an overtreatment price vector in *Salience* is higher than in *NoSalience* in all periods, with an increase in the size of the difference overtime. One possible explanation is that in *Salience*, the experts' costs become more salient, and hence an overtreatment price vector sends a stronger signal to customers that experts will choose a major treatment. Given a major treatment, a customer can get either a proper treatment (in case she has a major problem) or an overtreatment (in case she has a minor problem). In both cases, her payoff is $10 - \bar{p}$. Consider the possible values of $\bar{p}$ and $\underline{p}$ in our experiment and the fact that our sample is risk-averse (see Table 4.1), $10 - \bar{p}$ can be a more attractive payoff compared to the payoff of being undertreated (-$\underline{p}$) or the outside option (u = 1.6 ECU).

4.4.2 Prices and markups

Next, we analyze how the probability of posting undertreatment price vectors varies with salience. In our experiment, experts post undertreatment vectors frequently no matter the treatment. 79.2% of the price vectors under *NoSalience* and 77.4% of the price vectors in *Salience* are undertreatment vectors. At the aggregate level, the share of undertreatment vectors is not significantly different (p=0.649, t-test with clustering at subject level).

At the individual level, Table 4.3 reveals, however, that, keeping everything else constant, experts are significantly less likely to post an undertreatment vector under *Salience* than *NoSalience*. We find that the probability of posting an undertreatment price vector in *Salience* is 17 percentage points lower compared to *NoSalience*. Our finding goes in line with Hypothesis 2, namely, that experts are more likely to post undertreatment price tariffs in *NoSalience* than in *Salience*.

The impact of salience on the likelihood of posting an undertreatment vector varies for experts with different social preferences. Selfish experts are 20 percentage points more likely to post undertreatment tariffs in *NoSalience* than in *Salience*, while for pro-social experts the likelihood does not change significantly across treatments. The number of pro-social experts in the market does not seem to matter for price-setting behavior.

TABLE 4.3: Probability of posting an undertreatment vector

Undertreatment vector	All experts	Pro-social experts	Selfish experts
Salience	-0.17***	-0.05	-0.20**
	(0.05)	(0.09)	(0.08)
Pro-social market		0.03	0.04
		(0.04)	(0.05)
Major price	☑	☑	☑
Undertreatment vector $_{t-1}$	☑	☑	☑
Interaction $_{t-1}$	☑	☑	☑
Sufficient $_{t-1}$	☑	☑	☑
Individual controls	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
N	900	360	540

Notes: Probit estimations, average marginal effects are displayed. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. The subscript $_{t-1}$ indicates values from the immediate previous period. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs.

Regarding markup differences, our result suggests that the markup difference is, on average, negative in both treatments with mean values of -1.91 and -1.47 in *NoSalience* and *Salience*, respectively. Figure 4.4 shows that the average markup difference in *Salience* is less negative than in *NoSalience*, i.e., prices set are significantly closer to the equal markup prices predicted by standard theory (Dulleck and Kerschbamer, 2006). This difference is significant at the aggregate level (p=0.027, t-test with clustering at subject level).

Additionally, Table 4.4 shows a substantial effect of *Salience* at an individual level: Experts in *Salience* post price vectors with a significantly higher markup difference, implying a markup difference that is closer to zero. We account for market characteristics and observe that the markup difference is also heavily affected by the overall price level and inertia, and we include them in our regression analysis to control for these possible explanations. However, experts do not seem to account for interaction in the previous period and sufficiency of the treatment they have previously provided.

Social preference classification provides rather surprising evidence: salience has the opposite impact on the average markup difference for pro-social and selfish experts. We find that the increase in the markup difference we have documented at the aggregate and market levels is driven entirely by selfish experts. Salience has a substantial positive effect for them. Pro-social experts, on the

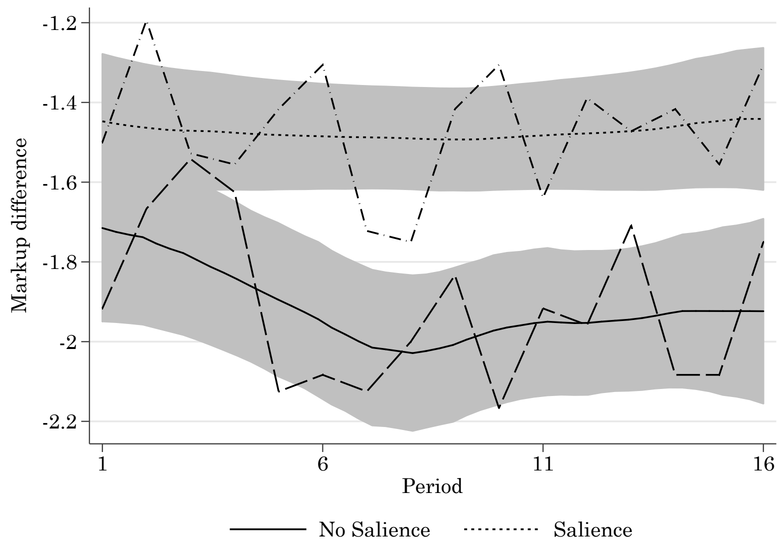

FIGURE 4.4: **Average markup difference**

Note: The figure displays kernel density estimations of average markup differences over time. Fitted values are estimated using Epanechniov kernel with optimal bandwidth. Gray areas correspond to 95% confidence intervals.

contrary, post price vectors with an even lower markup difference in *Salience* in comparison to *NoSalience*, implying a tendency to be further away from an equal markup.

Is the increase in the markup difference under *Salience* driven by an increase of $\overline{p}$, a decrease of $\underline{p}$, or both? We observe that the higher markup difference in *Salience* is driven mainly by the lower price of the minor treatment. The price of the minor treatment is, on average, 5.68 in *NoSalience*, and 5.39 in *Salience*. The price of the major treatment is, on average, 7.77 in *NoSalience* and 7.93 in *Salience* (see Table 4.10).

Regarding the mechanism of price setting for pro-social and selfish experts, there is no effect of salience on the price of the major treatment: neither for pro-social, nor for selfish experts. Instead, the difference in their behavior is captured entirely by $\underline{p}$. Pro-social experts set the price of the minor treatment 0.8 ECU higher in *Salience* compared to *NoSalience*, while selfish experts, on the contrary, lower it by 1.2 ECU in the *Salience*. The price of the minor treatment therefore drives the gap of the salience effect on the markup difference of about 2 ECU (see Table 4.10).

One counterargument could be that treatment *Salience* reduces the overall complexity of the experiment. However, our data show that time trends in both prices are very similar in *Salience* and *NoSalience*, which suggests that price differences can be explained by treatment and not by difference in learning (see Figure 4.5). Parallel development of prices over 16 periods helps us to rule out this potential explanation.

TABLE 4.4: Markup difference

Markup difference Δ	All experts			Pro-social	Selfish
	(1)	(2)	(3)	experts	experts
Salience	2.93***	1.22***	1.22***	-0.78**	1.76***
	(0.34)	(0.26)	(0.26)	(0.31)	(0.50)
Pro-social market				-0.23	-0.26
				(1.37)	(0.29)
Major price		0.55***	0.55***	0.67***	0.43***
		(0.06)	(0.06)	(0.08)	(0.08)
Markup difference $_{t-1}$		0.52***	0.52***	0.19***	0.41***
		(0.04)	(0.04)	(0.07)	(0.06)
Interaction $_{t-1}$			0.01	-0.17	0.12
			(0.13)	(0.13)	(0.19)
Sufficient $_{t-1}$			-0.05	-0.13	-0.11
			(0.14)	(0.16)	(0.20)
Individual controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
N	960	900	900	360	540

Notes: OLS estimations of markup differences. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. The subscript $_{t-1}$ indicates values from the immediate previous period. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences and beliefs.

4.4.3 Mistreatment

If a customer decides to interact upon posted prices, an expert observes the severity of her problem and chooses which treatment to provide. Given verifiability, there is no scope for overcharging, i.e., experts cannot provide one type of treatment and charge customers for another type. However, experts may still mistreat customers. Mistreatment can generally occur in two cases: when a customer with a minor problem receives a major treatment (overtreatment) and when a customer with a major problem receives a minor treatment (undertreatment). Under- and overtreatment rates are calculated as a share of all under-/overtreatments given under-/overtreatment was possible (i.e., undertreatment rates only for customers with a major problem, overtreatment rates only for customers with a minor problem).

Undertreatment rates are 53.66% and 49.69% in *NoSalience* and *Salience*, respectively. Overtreatment rates are 20.19% and 20.57% in *NoSalience* and *Salience*, respectively. We estimate the probability of sufficient treatment provision for customers with a major problem and show the results in Table 4.5. We find that, overall, customers in *Salience* are more likely to receive a sufficient treatment in comparison to those in *NoSalience*.

The impact of salience is insignificant for selfish experts, but large and highly significant for pro-social ones: Pro-social experts are almost twice as likely to provide a sufficient treatment in *Salience* compared to *NoSalience*. Interestingly, despite posting more undertreatment price vectors, pro-social experts undertreat less. Selfish experts, on the contrary, post fewer undertreatment vectors in *Salience*, however, their likelihood of sufficient treatment provision does not vary significantly.

TABLE 4.5: Probability of sufficient treatment

Sufficient	All experts			Pro-social experts	Selfish experts
	(1)	(2)	(3)		
Salience	0.34***	0.27**	0.23*	0.97***	0.40
	(0.12)	(0.13)	(0.13)	(0.36)	(0.25)
Pro-social market				-0.16	-0.15
				(0.16)	(0.15)
Undertreatment vector		☑	☑	☑	☑
Overtreatment vector			☑	☑	☑
Individual controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
N	241	241	241	86	155

Notes: Probit estimations, average marginal effects are displayed. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs. We also include time (period) and market fixed effects.

Generally, experts in *Salience* make more efficient decisions. The probability of providing a sufficient treatment conditional on a customer having a minor problem is 50.31%, and it differs a lot depending on the price vector chosen by the expert in this period. It is less likely an expert provides a sufficient treatment if he posts an undertreatment vector (41.60%) and more likely otherwise

(82.35%). In *NoSalience*, this pattern is much less pronounced. Probability of providing a sufficient treatment conditional on a customer having a minor problem is 46.34%, and it differs rather little depending on the price vector set: in case an undertreatment vector has been posted, an expert provides a sufficient treatment with the probability of 45.07%. Otherwise, the probability of sufficient treatment provision is higher (54.55%) but only marginally.

TABLE 4.6: Probability of overtreatment

Overtreatment	All experts			Pro-social experts	Selfish experts
	(1)	(2)	(3)		
Salience	0.50***	0.20**	0.14*	0.19	0.38***
	(0.13)	(0.09)	(0.08)	(0.13)	(0.15)
Pro-social market				-0.15**	-0.22**
				(0.07)	(0.10)
Undertreatment vector		☑	☑	☑	☑
Overtreatment vector			☑	☑	☑
Individual controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
N	241	241	241	74	158

Notes: Probit estimations, average marginal effects are displayed. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. The subscript $_{t-1}$ indicates values from the immediate previous period. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs. We also include time (period) and market fixed effects.

Average overtreatment probabilities are very similar in *Salience* (20.57%) and *NoSalience* (20.19%). However, we find that, conditional on other market outcomes, customers in *Salience* are more likely to be overtreated than in *NoSalience* (see Table 4.6). More precisely, if an expert in *Salience* posts an overtreatment vector, and a matched customer has a minor problem, the expert overtreats with certainty (the probability of 100%), whereas the probability of overtreatment is only 7.44% if another price vector was posted. In *NoSalience* the pattern is similar but less pronounced. If an overtreatment price vector was posted, customers are 55.56% likely to be overtreated, and 16.84% otherwise. Social expert classification shows that the probability of overtreatment increases with salience but only for selfish experts. Pro-social experts have an insignificant increase in their likelihood of overtreating, whereas selfish experts are about 38 percentage points more likely to overtreat in *Salience*.

4.4.4 Welfare and efficiency

Various key market outcomes, such as markup differences and mistreatment rates, are influenced by salience, which can lead to welfare implications. We analyze how welfare differs between treatments, and break it down to customer and producer surplus in Table 4.7.

In addition, we construct a market-level efficiency (as Mimra et al. (2016) do). We calculate the efficiency level as cumulative profits in the market less the outside option of all players, and normalize it with respect to the distribution of customers in the respective market, which allows us to account for the random differences in total welfare generated by the severity of the customers' problems.[14] The analysis on efficiency is shown in Table 4.8.

We start by analyzing welfare through profits acquired by participants over the course of the experiment, and analyze them at the treatment and market level. Table 4.7 shows that although salience leads to an improvement in a number of market outcomes, cumulative profits go down. Moreover, the loss is driven entirely by the loss in customer surplus. In contrast, experts benefit greatly from their profits being displayed to customers.

We observe several patterns in total welfare. Total welfare decreases if there is interaction unless it is sufficient because customers experience a large instant loss from undertreatment. Total welfare remains unchanged if the markup difference increases through $\overline{p}$: In this case, customers lose, on average, the same amount that experts gain. However, if the markup difference increase comes through the reduction of $\underline{p}$, customer surplus does not change significantly, whereas producer surplus decreases, so total welfare goes down as well.[15]

However, our data also show that, despite the theoretical probability of customers to have a minor issue is 50%, minor issue actually arises in 47% and 56% of cases in *Salience* and *NoSalience*, respectively. As mentioned above, the severity of the treatment crucially affects the cumulative profits a customer-expert pair can generate and thus, it is important to take it into account for estimating efficiency.

[14]Given interaction, every consumer is randomly assigned to have a major or a minor issue with a probability of 50%. In case of a minor issue, every customer-expert pair can generate at least $(10 - p) + (p - 6) = 4$ and at most $(10 - p) + (p - 2) = 8$. In case of a major issue, every customer-expert pair can only generate $(0 - p) + (p - 2) = -2$ in the worst case and $(10 - p) + (p - 6) = 4$ in the best case. We thus account for these differences while calculating the market efficiency measure.

[15]Table 4.11 breaks down consumer surplus and producer surplus for pro-social and selfish types.

TABLE 4.7: Welfare

Welfare	Per treatment			Per market		
	CS	PS	TS	CS	PS	TS
Salience	-9.38***	4.89***	-4.50***	-1.31***	0.82***	-0.58**
	(0.57)	(0.34)	(0.49)	(0.34)	(0.19)	(0.23)
Major price	0.10	0.27*	0.37*	-0.38***	0.63***	0.30***
	(0.26)	(0.15)	(0.20)	(0.14)	(0.08)	(0.10)
Markup difference	-0.16	-0.22**	-0.38**	-0.05	-0.27***	-0.30***
	(0.17)	(0.10)	(0.15)	(0.12)	(0.05)	(0.08)
Interaction	-2.90***	0.73	-2.17***	-6.66***	2.17***	-4.46***
	(0.76)	(0.47)	(0.72)	(0.48)	(0.25)	(0.33)
Sufficient	4.26***	-0.48	3.78***	7.95***	-0.55**	7.47***
	(0.81)	(0.50)	(0.74)	(0.49)	(0.27)	(0.34)
Individual controls	No	No	No	Yes	Yes	Yes
Fixed effects	No	No	No	Yes	Yes	Yes
N	1920	1920	1920	960	960	1920

Notes: OLS estimations, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs. We also include time (period) and market fixed effects.

TABLE 4.8: Efficiency

Efficiency	All	Pro-social	Selfish
Salience	0.05***	0.02**	0.11***
	(0.01)	(0.01)	(0.01)
Pro-social market	-0.06***	-0.04***	-0.08***
	(0.00)	(0.00)	(0.00)
Major price	☑	☑	☑
Markup difference	☑	☑	☑
Interaction	☑	☑	☑
Sufficient	☑	☑	☑
Individual controls	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
N	1920	896	1024

Notes: OLS estimations, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs. We also include time (period) and market fixed effects.

We find that efficiency indeed increases under *Salience*. On average, efficiency increases by 5 percentage points with salience. The effect is significant for subsamples with different social preferences. However, we find especially pronounced efficiency gain from salience for selfish subjects: It accounts for 11 percentage points increase in market efficiency on average.

Overall, market efficiency increases, selfish individuals benefit from salience, whereas pro-social individuals on average do not. Salience particularly harms pro-social customers but especially benefits pro-social experts, whereas selfish customers remain relatively unaffected, and selfish experts benefit but to a smaller extent.

4.5 Conclusion

There exist contradictions between theoretical predictions and empirical evidence on the role of verifiability in the credence goods market. While theory predicts that under certain conditions, verifiability leads to market efficiency, observations from real markets go against this prediction. We are the first to provide theoretical argument and experimental evidence that customer's limited attention plays a role in this inconsistency. Our finding goes in line with recent advocacy for more transparency on experts' pay in credence goods markets such as healthcare or repair services.

Based on the inherent features of lab experiments on credence goods market, we set up a model of a monopolistic credence goods market where customers pay limited attention to expert's costs, and full attention to prices and valuation of a proper treatment. Our model further assumes that experts know customers pay limited attention to their costs while customers are unaware thereof. Our main hypotheses are that an increase in customers' attention with regard to experts' costs results in (i) a decrease in customer interaction given an undertreatment tariff, (ii) a decrease in the amount of undertreatment tariffs and insufficient treatments, and (iii) a smaller markup difference between the major treatment and the minor treatment.

We test the hypotheses in a laboratory experiment, and confirm the last two hypotheses. We observe less undertreatment and experts' price vectors were significantly closer to equal mark-up pricing when expert costs are made salient than when they are not. We do not find strong supporting evidence for the first hypothesis. Interestingly, we observe that interaction given an overtreatment tariff under the salience of experts' cost is much higher than under no salience. We argue that risk aversion and experimental parameterization might account for this effect. In terms of welfare, the salience of experts' costs leads to an increase in accumulated payoffs. Throughout, we observe a heterogeneity of results with regard to social preference.

Overall, our results suggest that customers' limited attention is a possible explanation for the empirical evidence on the inefficiency of verifiability in credence goods markets. Furthermore, our study draws a rather nuanced picture when it comes to the merits of introducing more transparency of experts' costs. We observe a positive effect on undertreatment, markups, and welfare but we do not find an overall increase in interaction compared to the case without transparency. Hence, increasing transparency might serve customers who choose to interact and all experts, but might do more harm than good to customers who interact less, or refrain from interaction altogether. Taken on its own, our findings explain why providers in healthcare and repair service appear to not object to calls for more transparency. What remains an open question for future research is whether expert providers aim to gain a competitive advantage over their rivals through transparency.

4.A Proofs

We differentiate among three cases: $\beta = 1$, $0 < \beta < 1$, and $\beta = 0$.

Case 1: $\beta = 1$

This is when customers are rational and equally attentive to all features. As shown by Dulleck and Kerschbamer, 2006, in equilibrium we have:

- The expert posts equal-markup prices:

$$\bar{p} = v + (1-h)(\bar{c} - \underline{c}) - u$$
$$\underline{p} = v - h(\bar{c} - \underline{c}) - u.$$

- The expert provides sufficient conditions.
- The customer chooses to interact.

Case 2: $0 < \beta < 1$

When the customer is inattentive to the hidden feature, she expects the expert to post the following prices in equilibrium:

$$\bar{p}_c = v + \beta(1-h)(\bar{c} - \underline{c}) - u$$
$$\underline{p}_c = v - \beta h(\bar{c} - \underline{c}) - u.$$

Consider three classes of tariffs perceived by the customer:

(i) The markup for the major treatment exceeds that for the minor treatment ($\bar{p} - \beta\bar{c} > \underline{p} - \beta\underline{c}$),

(ii) the markup of the minor treatment exceeds that for the major treatment ($\bar{p} - \beta\bar{c} < \underline{p} - \beta\underline{c}$), and

(iii) markups are the same for both treatments ($\bar{p} - \beta\bar{c} = \underline{p} - \beta\underline{c}$).

A customer with limited attention expects the following:

- The expert performs the major treatment if he posts (i), he performs the minor treatment if he posts (ii), and he is indifferent if he posts (iii).[16] The customers observe the price and infer the experts' incentives accordingly.

[16]Similarly to Dulleck and Kerschbamer, 2006, we assume that the expert is indifferent between two treatments if he posts an equal-markup tariff. Moreover, this is common knowledge.

- The highest obtainable profit in each case amounts to:
  (i) $v - u - \beta\bar{c}$
  (ii) $(1-h)v - u - \beta\underline{c}$
  (iii) $v - u - \beta(h\bar{c} + (1-h)\underline{c})$.

Given that $u > 0, \bar{c} > \underline{c}, v > (\bar{c} - \underline{c}), \beta \in (0,1)$, and $h \in [0,1]$, the equal-markup tariff gives the highest obtainable profit for the experts.

Proof. **(Lemma 1)**
$\forall \beta \in (0,1)$, $\underline{p}_c$ is strictly larger than $\underline{p}$ and $\bar{p}_c$ is strictly smaller than $\bar{p}$. □

Proof. **(Proposition 1)**
We denote the markup difference given $(\underline{p}, \bar{p})$ as Δ^* and the markup difference given $(\underline{p}_c, \bar{p}_c)$ as Δ^*_c. It is clear from Lemma 1 that $\Delta^*_c < \Delta^*$. Furthermore, we denote tariffs where the markup for the minor treatment exceeds that for the major treatment as U, tariffs where the markup for the major treatment exceeds that for the minor treatment as O, and equal markup tariffs as E.
When an expert posts a tariff, the markup difference of his tariff, denoted as Δ can fall into one of five cases: $\Delta < \Delta^*_c$, $\Delta = \Delta^*_c$, $\Delta^*_c < \Delta < \Delta^*$, $\Delta = \Delta^*$, and $\Delta > \Delta^*$. Table 4.9 provides an overview of the true tariff, the tariff perceived by customers with limited attention, and the treatment chosen by the expert in each case.

TABLE 4.9: Overview of tariffs and chosen treatment

	$\Delta < \Delta^*_c$	$\Delta = \Delta^*_c$	$\Delta^*_c < \Delta < \Delta^*$	$\Delta = \Delta^*$	$\Delta > \Delta^*$
True tariff	U	U	U	E	O
Tariff perceived by customers	U	E	O	O	O
Chosen treatment	Minor	Minor	Minor	Minor/Major	Major

Notes: The table summarizes five cases that a specific tariff can possibly belong to. For each case, the table shows the category of the true tariff, of the tariff perceived by customers with limited attention, and the treatment chosen by experts. Abbreviations: U: tariffs where the markup for the minor treatment exceeds that for the major treatment. O: tariffs where the markup for the major treatment exceeds that for the minor treatment. E: equal markup tariffs.

□

Case 2.1: $\Delta < \Delta_c^*$

Any tariffs in this case are perceived by customers as having a markup for the minor treatment exceeding that for the major treatment, and it is really the case. Therefore, the expert will always choose to provide the minor treatment. The highest price accepted by the customer for the minor treatment is $(1-h)v-u$. Thus, the highest obtainable profit of the expert given this price is:

$$\Pi_1 = (1-h)v - u - \underline{c}$$

Case 2.2: $\Delta = \Delta_c^*$

The tariff in this case is $(\underline{p}_c, \bar{p}_c)$, which is perceived as an equal markup tariff by the customers. In fact, this is a tariff where the markup of the minor treatment exceeds that for the major treatment. Thus, the expert will always provide the minor treatment, and will be able to attract all customers in the market to interact with him.

The highest obtainable profit of the expert is:

$$\Pi_2 = \underline{p}_c - \underline{c} = v - \beta h(\bar{c} - \underline{c}) - u - \underline{c}$$

Case 2.3: $\Delta_c^* < \Delta < \Delta^*$

Any tariffs in this case are perceived by the customer as having a markup for the major treatment exceeds that for the minor treatment. In fact, the opposite is true. These are tariffs where the markup for the minor treatment exceeds that for the major one. Thus, the expert will always provide the minor treatment, and will be able to attract all customers in the market to interact with him.

The highest obtainable profit of the expert given this price is:

$$\Pi_3 = v - u - \underline{c}$$

Case 2.4: $\Delta = \Delta^*$

The tariff in this case is the equal markup tariff $(\bar{p}, \underline{p})$. The expert will choose the minor treatment for a minor issue, and the major treatment for a major issue.

The customers will perceive it as a tariff where the markup for the major treatment exceeds that for the minor treatment. Thus, the customers anticipate that the expert will always choose to provide the major treatment and charge price $\bar{p}$.

If the expert chooses the major treatment, his highest obtainable profit is:

$$\bar{\Pi}_4 = \bar{p} - \bar{c} = v - u - h\bar{c} + (h-1)\underline{c}.$$

If the expert chooses the minor treatment, his highest obtainable profit is:

$$\underline{\Pi}_4 = \underline{p} - \underline{c} = v - u - h\bar{c} + (h-1)\underline{c}$$

Case 2.5: $\Delta > \Delta^*$

Any tariffs in this case are perceived by the customer as having the markup for the major treatment exceeds that for the minor treatment, and this really is the case. Therefore, the expert will always choose to provide the major treatment. The highest price accepted by the customer for the major treatment is $v - u$. Therefore, the highest obtainable profit of the expert is:

$$\Pi_5 = v - u - \bar{c}$$

Case 3: $\beta = 0$

When customers completely neglect the hidden feature, experts post a tariff where the price for the major treatment is equal to that for the minor treatment. Any tariffs in this case are perceived as an equal markup tariff by the customers. In fact, this is a tariff where the markup of the minor treatment exceeds that for the major treatment. Thus, the expert will always provide the minor treatment, and will be able to attract all customers in the market to interact with him.
The highest obtainable profit of the expert given this price is:

$$\Pi_6 = \underline{p} - \underline{c} = v - u - \underline{c}$$

Given that $u > 0, \bar{c} > \underline{c}, v > (\bar{c} - \underline{c}), \beta \in [0,1)$ and $h \in [0,1]$, it is clear that $\Pi_1 < \Pi_2$, $\Pi_4 < \Pi_2$, $\Pi_5 < \Pi_2$, $\Pi_2 < \Pi_3$, and $\Pi_3 = \Pi_6$. Therefore, conditional on interaction, the expert will always post tariffs which satisfy $\Delta_c^* < \Delta < \Delta^*$ and will always choose to provide the minor treatment.

4.B Additional results

TABLE 4.10: Prices

Prices	All experts		Pro-social experts		Selfish experts	
	$\underline{p}$	$\overline{p}$	$\underline{p}$	$\overline{p}$	$\underline{p}$	$\overline{p}$
$Salience$	-0.93***	0.12	0.81***	0.25	-1.23**	0.23
	(0.26)	(0.22)	(0.28)	(0.25)	(0.48)	(0.38)
Pro-social market			0.44	0.82	0.16	0.23
			(1.61)	(1.78)	(0.26)	(0.22)
Minor price $_{t-1}$	0.54***		0.33***		0.57***	
	(0.05)		(0.07)		(0.07)	
Major price $_{t-1}$		0.47***		0.43***		0.43***
		(0.05)		(0.08)		(0.07)
Markup difference $_{t-1}$	-0.09**	0.03	-0.00	-0.13**	0.05	-0.01
	(0.04)	(0.03)	(0.07)	(0.07)	(0.06)	(0.05)
Interaction $_{t-1}$	0.40***	0.30***	0.35**	0.23	0.39**	0.30**
	(0.11)	(0.11)	(0.14)	(0.15)	(0.16)	(0.14)
Sufficient $_{t-1}$	0.05	0.03	0.19	0.17	0.06	-0.05
	(0.12)	(0.12)	(0.16)	(0.18)	(0.18)	(0.15)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	900	900	360	360	540	540

Notes: OLS estimations of prices. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. The subscript $_{t-1}$ indicates values from the immediate previous period. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs.

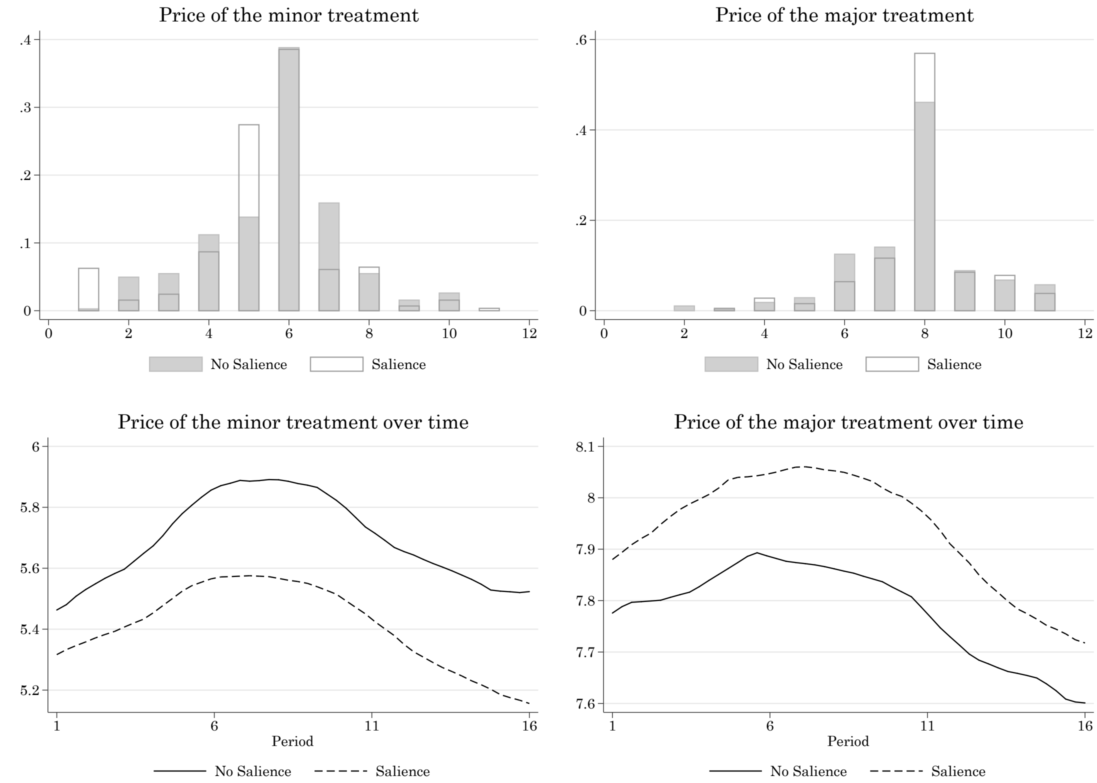

FIGURE 4.5: **Distribution of prices**

Note: The figure displays distributions of the price for minor treatment (upper left) and major treatment (upper right) and kernel density estimations of price for minor treatment (lower left) and major treatment (lower right) over time. Fitted values are estimated using the Epanechniov kernel with optimal bandwidth.

TABLE 4.11: Welfare by social preference classification

Welfare	Customer surplus		Producer surplus		Total surplus	
	Pro-social	Selfish	Pro-social	Selfish	Pro-social	Selfish
Salience	-1.54***	-0.92	1.46***	1.06**	-0.27	0.88**
	(0.57)	(0.63)	(0.36)	(0.31)	(0.35)	(0.40)
Pro-social market	0.06	-0.21	-0.49*	-0.55***	-0.37**	-1.31***
	(0.24)	(0.39)	(0.28)	(0.13)	(0.18)	(0.20)
Major price	☑	☑	☑	☑	☑	☑
Markup difference	☑	☑	☑	☑	☑	☑
Interaction	☑	☑	☑	☑	☑	☑
Sufficient	☑	☑	☑	☑	☑	☑
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	512	448	384	576	896	1024

Notes: OLS estimations, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. The subscript $_{t-1}$ indicates values from the immediate previous period. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and beliefs. We also include time (period) and market fixed effects.

4.C Experimental instructions

(Translated from German)

Thank you for your participation in this experiment. Please do not to talk to any other participants during the experiment. Today's experiment consists of several parts. Your earning is the total income from these parts. In addition, you will receive a show-up fee of 4 EUR for today's participation and for answering the questionnaire.

INSTRUCTIONS

2 Roles and 16 Rounds

This experiment consists of **16 rounds**, each of which consists of the same sequence of decisions. This sequence of decisions is explained in detail below.

There are 2 kinds of roles in this experiment: **player A** and **player B**. At the beginning of the experiment you will be randomly assigned to one of these two roles. On the first screen of the experiment you will see which role you are assigned to. Your role remains the same lthroughout the experiment. In your group there are 4 players A and 4 players B.

One player A always interacts with one player B. However, the pairs **change** after each round. That means you will interact with a **new** player (the other role) every round.

All participants get the same information on the rules of the game, including the costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 3 decisions which are made consecutively. Decisions 1 abd 3 are made by player A, decision 2 is made by player B.

1. Player A chooses one price for action 1 and one price for action 2.
2. Player B gets to know the prices chosen by player A. Then player B decides whether he/she wants to interact with player A. If not, this round ends for him/her.

If yes...

3. Player A (but **not** player B) is informed whether player B is of type 1 or type 2. Player A chooses thereupon either action 1 or action 2. Player B has to pay the price specified by player A in decision 1 for the action chosen by player A.

Detailed Illustration of the Decisions and Their Consequences Regarding Payoffs

Decision 1

- **Player A** has to choose between 2 actions (action 1 and action 2) at decision 3.
- **Action 1** costs player A **2 points** (= currency of the experiment).
- **Action 2** costs player A **6 points.**
- Player A can charge prices for these actions from player B who decide to interact with him/her. At **decision 1** each Player A has to **set the prices for both actions**. Only (strictly) positive integer numbers are possible, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 are valid prices.
- Note that the price for action 1 must not exceed the price for action 2.

Decision 2
(Instruction of Decision 2 differs between two treatments.)

(*NoSalience* treatment)

- **Player B** gets to know the **prices** of player A for the two actions at decision 1. Then player B decides whether he/she wants to interact with player A or not.
- **If he/she wants to do so**, player A can choose an action at decision 3 and charge a price for that action (see below). **If he/she doesn't want to interact,** this round **ends** for player B and he/she gets a **payoff of 1.6 points** for this round.

(*Salience* treatment)

- **Player B** gets to know the **prices and profits** of player A for the two actions at decision 1. Then player B decides whether he/she wants to interact with player A or not.
- **If he/she wants to do so**, player A can choose an action at decision 3 and charge a price for that action (see below). **If he/she doesn't want to interact,** this round **ends** for player B and he/she gets a **payoff of 1.6 points** for this round.

Decision 3

- Before decision 3 is made (in case player B choses "Yes" at decision 2) a type is randomly assigned to player B. **Player B** can be one of the two types: **type 1** or **type 2**. This type is randomly determined for each player B **in each new round**.
- With a probability of **50% player B is of type 1**, and with a probability of **50% he/she is of type 2**. Imagine that a coin is tossed for each player B in each round. If the result is e.g. "heads", player B is of type 1, if the result is "tails" he/she is of type 2.
- Every **player A** gets to **know the types of player B** who interact with him/her before he makes his decision 3. Then player A chooses an action for each player B, either action 1 or action 2.
- An **action** is **sufficient** in the following cases:

  a) Player B has type 1 and player A chooses either action 1 or action 2.

  b) Player B has type 2 and player A choose action 2.
- An action is **not sufficient**, if player B has type 2 but player A chooses action 1.
- **Player B** receives **10 points**, if the action chosen by player A is **sufficient**. **Player B** receives **0 point**, if the action chosen by player A is **not sufficient**.
- **At no time player B** will be informed whether he/she is of type 1 or a type 2 player in each round, as well as which action player A has chosen.
- **Player A** charges player B the **price** set out in decision 1 for the action chosen in decision 3.

Payoffs
If player B chose not to interact with any of the players A (*decision "No" from player B*), both player A and player B get **1.6 points** for this particular round.
Otherwise (*decision "Yes" by player B*) the payoffs are as follows:
Player A receives the according **price** (denoted in points) he/she set at decision 1 **less the costs** for the action chosen at decision 3.
The payoff of **player** B depends on whether the action chosen by player A in decision 3 was sufficient or not:
a) If the action chosen by player A was sufficient, Player B gets 10 points less the price set in decision 1 for the action chosen at decision 3.

b) If the action chosen by player A was not sufficient, Player B has to pay the price set in decision 1 for the action chosen at decision 3.
At the beginning of the experiment you receive an **initial endowment of 6 points**. With this endowment you are able to cover losses that might occur in some rounds. Losses can also be compensated by gains in other rounds. If your total payoff sums up to a loss at the end of the experiment you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this term. **Please note that there is always a possibility to avoid losses in this experiment.**
To calculate the payoff of this part, the initial endowment and the profits of all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 25 Euro-cents
(i.e. 4 points = 1 EUR)

You will see all further instruction on the computer screen.

4.D Questionnaire

The questionnaire at the end of the experiment contains the following items (translated from German):

1. **Elicitation of beliefs:**
   (Only for sellers)
   When you set the price, did you expect that Player B will decide to interact? (Yes/No)
   Which action (Action 1 or Action 2) would you choose given the following scenarios?
   Price: 3 for Action 1 and 8 for Action 2
   Price: 4 for Action 1 and 8 for Action 2
   Price: 5 for Action 1 and 8 for Action 2
   Price: 6 for Action 1 and 8 for Action 2
   Price: 7 for Action 1 and 8 for Action 2
   (Only for buyers):
   As you decided to interact, did you expect that Player A will choose a sufficient action?
   Which action (Action 1 or Action 2) do you expect Player A to choose given the following scenarios? Price: 3 for Action 1 and 8 for Action 2

Price: 4 for Action 1 and 8 for Action 2
Price: 5 for Action 1 and 8 for Action 2
Price: 6 for Action 1 and 8 for Action 2
Price: 7 for Action 1 and 8 for Action 2

2. **Risk preference, general risk question:** same wording as in German Socio-Economic Panel questionnaire (SOEP, see, for example, Wagner et al., 2007)
How do you evaluate yourself? Are you generally a risk-seeking person or do you try to avoid risks? The leftmost box means "not at all risk-seeking" and the rightmost "very risk-seeking". With the boxes in between, you can graduate your statement.

not at all risk-seeking ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ *very risk-seeking*

3. **Risk preference, incentivized choice list:** Subjects make eleven, pairwise decisions between a lottery with a fifty-fifty chance of winning either 2 EUR or 7 EUR and a safe payment. The safe payment increases in 0.5 EUR increments, ranging from 2 EUR to 7 EUR.

4. **Loss aversion** similar to Karle et al., 2015
You will answer questions related to lotteries. If you accept the lotteries, you can make either a profit or a loss. Below are six different lotteries. For each lottery, you can decide whether to accept or to reject it. If you reject, your payment remains unchanged. If you accept, your earning will make either an additional profit or an additional loss.

At the end of the experiment, one of the six lotteries will be randomly selected. So you should make every decision as if it were your only decision. The selected lottery is then randomly drawn to determine whether the additional profit or loss will be realized for you.

(All with the same options: Accept or Reject)

Lottery 1: With a 50% probability you lose 2 EUR and with a 50% probability you win 6 EUR.

Lottery 2: With a 50% probability you lose 3 EUR and with a 50% probability you win 6 EUR.

Lottery 3: With a 50% probability you lose 4 EUR and with a 50% probability you win 6 EUR.

Lottery 4: With a 50% probability you lose 5 EUR and with a 50% probability you win 6 EUR.

Lottery 5: With a 50% probability you lose 6 EUR and with a 50% probability you win 6 EUR.

Lottery 6: With a 50% probability you lose 7 EUR and with a 50% probability you win 6 EUR.

5. **Social preference** (survey question, Falk et al., 2018)
   Question 1: Imagine the following situation: Today you unexpectedly received 1000 EUR. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed).
   Question 2: Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize that you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about 20 EUR in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 EUR, the most expensive one costs 30 EUR. Do you give one of the presents to the stranger as a "thank you" gift?
   Which present do you give to the stranger?
   1. No, would not give present
   2. The present worth 5 EUR
   3. The present worth 10 EUR
   4. The present worth 15 EUR
   5. The present worth 20 EUR
   6. The present worth 25 EUR
   7. The present worth 30 EUR

6. **Description of reasoning for decisions**
   (Only for sellers)
   Please answer the following questions:
   How did you decide for the prices? Please describe what you thought when you set the prices.

   How did you decide for the actions? Please describe, what you thought when you choose the action.

   Did you change your strategy across periods? When yes, why?

   (Only for buyers)
   Please describe your thought when you made the decision whether or not

to interact.
Did you change your strategy across periods? When yes, why?

7. **Socio-demographics:** age, gender, final grade point average at academic high school, last math grade at academic high school, field of study, monthly disposable amount of money, political orientation, number of experiments already participated in the same lab.

Declaration of contribution

Hereby I, Chi Trieu, declare that this chapter, entitled "Consumers with limited attention in the credence goods market" is co-authored with Vasilisa Petrischeva, Alexander Rasch, and Christian Waibel. The contributions of each author to this paper is as follows:

- The conception of the idea was done by Vasilisa Petrischeva and Chi Trieu, and revised by Alexander Rasch and Christian Waibel.
- The experimental design was done by Vasilisa Petrischeva and Chi Trieu, with support from Alexander Rasch and Christian Waibel.
- Developing the theoretical framework was done by Chi Trieu, with support from the other authors.
- Programming the Z-tree codes was done by Chi Trieu, with support from Vasilisa Petrischeva.
- Conducting the experiment was done by Vasilisa Petrischeva and Chi Trieu.
- The empirical analysis and graphical representation of the results was done by Vasilisa Petrischeva, with substantial support from Christian Waibel and minor support from the other authors.
- All authors jointly and equally contributed to the writing of the manuscript.

Signatures of the co-authors:

Vasilisa Petrischeva

Alexander Rasch

Christian Waibel

Bibliography

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). “Reference Points and Effort Provision”. *American Economic Review* 101.2, pp. 470–92.

Ahern, K. and A. K. Dittmar (2012). “The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation”. *The Quarterly Journal of Economics* 127.1, pp. 137–197.

Almås, I., A. Cappelen, K. G. Salvanes, E. Sørensen, and B. Tungodden (2016). “Willingness to Compete: Family Matters”. *Management Science* 62.8, 2149–2162.

Araujo, F. A., E. Carbone, L. Conell-Price, M. W. Dunietz, A. Jaroszewicz, R. Landsman, D. Lamé, L. Vesterlund, S. W. Wang, and A. J. Wilson (2016). “The slider task: an example of restricted inference on incentive effects”. *Journal of the Economic Science Association* 2.1, pp. 1–12.

Arcidiacono, P. and M. Lovenheim (2016). “Affirmative Action and the Quality-Fit Trade-Off”. *Journal of Economic Literature* 54.1, pp. 3–51.

Armstrong, M. and J. Vickers (2012). “Consumer Protection and Contingent Charges”. *Journal of Economic Literature* 50.2, pp. 477–93.

automotorundsport.de (2020). *Transparenz bei Inspektion und Reparatur (Transparency in inspection and repair)*. Retrieved from https://www.auto-motor-und-sport.de/tech-zukunft/werkstatt/opel-mydigitalservice-transparenz-inspektion-reparatur/.

Bagde, S., D. Epple, and L. Taylor (2016). “Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India”. *American Economic Review* 106.6, pp. 1495–1521.

Balafoutas, L., A. Beck, R. Kerschbamer, and M. Sutter (2013). “What drives taxi drivers? A field experiment on fraud in a market for credence goods”. *The Review of Economic Studies* 80.3, pp. 876–891.

Balafoutas, L., B. J. Davis, and M. Sutter (2016). “Affirmative action or just discrimination? A study on the endogenous emergence of quotas”. *Journal of Economic Behavior and Organization* 127, pp. 87–98.

Balafoutas, L., R. Kerschbamer, and M. Sutter (Oct. 2015). “Second-Degree Moral Hazard in a Real-World Credence Goods Market”. *The Economic Journal* 127.599, pp. 1–18.

Balafoutas, L. and M. Sutter (2012). "Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory". *Science* 335.6068, pp. 579–582.

— (2019). "How uncertainty and ambiguity in tournaments affect gender differences in competitive behavior". *European Economic Review* 118.C, pp. 1–13.

Banerjee, R., N. D. Gupta, and M. C. Villeval (2018). "The spillover effects of affirmative action on competitiveness and unethical behavior". *European Economic Review* 101.C, pp. 567–604.

— (2020). "Feedback spillovers across tasks, self-confidence and competitiveness". *Games and Economic Behavior* 123, pp. 127 –170.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (Nov. 2009). "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124.4, pp. 1497–1540.

Beaurain, G. and D. Masclet (2016). "Does affirmative action reduce gender discrimination and enhance efficiency? New experimental evidence". *European Economic Review* 90, pp. 350 –362.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (June 2018). "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway". *The Review of Economic Studies* 86.1, pp. 191–239.

Bock, O., I. Baetge, and A. Nicklisch (2014). "hroot: Hamburg Registration and Organization Online Tool". *European Economic Review* 71.C, pp. 117–120.

Buser, T., M. Niederle, and H. Oosterbeek (May 2014). " Gender, Competitiveness, and Career Choices ". *The Quarterly Journal of Economics* 129.3, pp. 1409–1447.

Calsamiglia, C., J. Franke, and P. Rey-Biel (2013). "The incentive effects of affirmative action in a real-effort tournament". *Journal of Public Economics* 98, pp. 15–31.

Cappelen, A. W., A. D. Hole, E. Sørensen, and B. Tungodden (2007). "The Pluralism of Fairness Ideals: An Experimental Approach". *American Economic Review* 97.3, pp. 818–827.

Cappelen, A. W., E. Sørensen, and B. Tungodden (2010). "Responsibility for what? Fairness and individual responsibility". *European Economic Review* 54.3, pp. 429 –441.

Charness, G. and P. Kuhn (2007). "Does Pay Inequality Affect Worker Effort? Experimental Evidence". *Journal of Labor Economics* 25.4, pp. 693–723.

Chattopadhyay, R. and E. Duflo (2004). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India". *Econometrica* 72.5, 1409–1443.

Chen, Z. C., D. Ong, and R. M. Sheremeta (2015). "The gender difference in the value of winning". *Economics Letters* 137, pp. 226 –229.

Chowdhury, S. M., J. Y. Jeon, and A. Ramalingam (2016). "Identity and group conflict". *European Economic Review* 90, pp. 107 –121.

Cohn, A., E. Fehr, and L. Götte (2015). "Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment". *Management Science* 61.8, 1741–2011.

Corgnet, B., R. Hernán-González, and E. Schniter (2015). "Why real leisure really matters: incentive effects on real effort in the laboratory". *Experimental Economics* 18.2, pp. 284–301.

Cornaglia, F., M. Drouvelis, and P. Masella (2019). "Competition and the role of group identity". *Journal of Economic Behavior & Organization* 162, pp. 136–145.

Council of European Union (1997). *Council Directive 97/81/EC*. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31997L0081.

— (2000). *Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000L0078.

Croson, R. and U. Gneezy (2009). "Gender Differences in Preferences". *Journal of Economic Literature* 47.2, pp. 448–474.

Dargnies, M.-P. (2012). "Men Too Sometimes Shy Away from Competition: The Case of Team Competition". *Management Science* 58.11, pp. 1982–2000.

Deller, C. and T. Sandino (2020). "Effects of a Tournament Incentive Plan Incorporating Managerial Discretion in a Geographically Dispersed Organization". *Management Science* 66.2, 503–1004.

Disability Rights Commission (2007). *Code of practice (revised) for providers of post-16 education and related services*.

Dreber, A., E. von Essen, and E. Ranehill (2011). "Outrunning the gender gap - boys and girls compete equally". *Experimental Economics* 14.4, pp. 567–582.

— (2014). "Gender and competition in adolescence: task matters". *Experimental Economics* 17.1, pp. 154–172.

Dulleck, U. and R. Kerschbamer (2006). "On Doctor, Mechanics, and Computer Specialists: The Economics of Credence Goods". *Journal of Economic Literature* 44.1, pp. 5–42.

Dulleck, U., R. Kerschbamer, and M. Sutter (2011). "The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition". *American Economic Review* 101.2, pp. 526–555.

Emons, W. (1997). "Credence Goods and Fraudulent Experts". *RAND Journal of Economics* 28.1, pp. 107–119.

Emons, W. (2001). “Credence Goods Monopolists”. *International Journal of Industrial Organization* 19.3-4, pp. 375–389.

European Commission (2019). *2019 Report on equality between women and men in the EU*.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (May 2018). “Global Evidence on Economic Preferences”. *The Quarterly Journal of Economics* 133.4, pp. 1645–1692.

Falk, A., F. Kosse, I. Menrath, P. E. Verde, and J. Siegrist (2017). “Unfair Pay and Health”. *Management Science* 64.4, 1477–1973.

Fallucchi, F. and S. Quercia (2018). “Affirmative action and retaliation in experimental contests”. *Journal of Economic Behavior and Organization* 156, pp. 23 – 40.

Federal Bureau of Investigation (2011). *Financial Crimes Report to the Public: Fiscal Year 2010-2011*. United States Department of Justice.

Federal Law Gazette (2015). *Gesetz für die gleichberechtigte Teilhabe von Frauen und Männern an Führungspositionen in der Privatwirtschaft und im öffentlichen Dienst (Law on Equal Participation of Women and Men in Leadership Positions in the Private and Public Sector)*. Bundesgesetzblatt Jahrgang 2015 Teil I Nr.17 S. 642.

Federal Ministry of Health (2020). *Patientenquittung (Patient Receipt)*. Retrieved from https://www.bundesgesundheitsministerium.de/themen/praevention/patientenrechte/patientenquittung.html.

Fischbacher, U. (2007). “z-Tree: Zurich toolbox for ready-made economic experiments”. *Experimental Economics* 10.2, pp. 171–178.

Fish, S. (2000). “The Nifty Nine Arguments against Affirmative Action in Higher Education”. *The Journal of Blacks in Higher Education* 27, pp. 79–81.

Frederick, S. (2005). “Cognitive Reflection and Decision Making”. *Journal of Economic Perspectives* 19.4, pp. 25–42.

Fullinwider, R. (2011). “Affirmative action”.

Gabaix, X. and D. Laibson (2006). “Shrouded attributes, consumer myopia, and information suppression in competitive markets”. *The Quarterly Journal of Economics* 121.2, 505–540.

Gächter, S., L. Huang, and M. Sefton (2016). “Combining “real effort” with induced effort costs: the ball-catching task”. *Experimental Economics* 19.4, pp. 687–712.

Gee, L. K. (2019). “The More You Know: Information Effects on Job Application Rates in a Large Field Experiment”. *Management Science* 65.5, pp. 2077–2094.

Gill, D. and V. Prowse (2012). “A Structural Analysis of Disappointment Aversion in a Real Effort Competition”. *American Economic Review* 102.1, pp. 469–503.

Gillen, B., E. Snowberg, and L. Yariv (2019). "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study". *Journal of Political Economy* 127.4, pp. 1826–1863.

Gneezy, U., M. Niederle, and A. Rustichini (Aug. 2003). "Performance in Competitive Environments: Gender Differences*". *The Quarterly Journal of Economics* 118.3, pp. 1049–1074.

Goerg, S. J., S. Kube, and J. Radbruch (2019). "The effectiveness of incentive schemes in the presence of implicit effort costs". *Management Science* 65.9, 3949–4450.

Gottschalk, F., W. Mimra, and C. Waibel (Feb. 2020). "Health Services as Credence Goods: a Field Experiment". *The Economic Journal* 130.629, pp. 1346–1383.

Greiner, B. (2004). *The Online Recruitment System ORSEE - A Guide for the Organization of Experiments in Economics*. Papers on Strategic Interaction. Max Planck Institute of Economics, Strategic Interaction Group.

Grubb, M. (2015a). "Failing to Choose the Best Price: Theory, Evidence, and Policy". *Review of Industrial Organization* 47.3, pp. 303–340.

Grubb, M. D. (2015b). "Consumer Inattention and Bill-Shock Regulation". *Review of Economic Studies* 82.1, pp. 219–257.

Gupta, N. D., P. Anders, and M. C. Villeval (2013). "Gender Matching and Competitiveness: Experimental Evidence". *Economic Inquiry* 51.1, pp. 816–835.

Heidhues, P. and B. Kőszegi (2017). "Naïveté-Based Discrimination". *The Quarterly Journal of Economics* 132.2, pp. 1019–1054.

— (2018). *Behavioral industrial organization*. Bernheim, D., DellaVigna, S., and Laibson, D., editors, Elsevier.

Heilman, M., C. Block, and J. Lucas (Aug. 1992). "Presumed Incompetent? Stigmatization and Affirmative Action Efforts". *Journal of Applied Psychology* 77.4, pp. 536–544.

Heilman, M. E., C. J. Block, and P. Stathatos (1997). "The Affirmative Action Stigma of Incompetence: Effects of Performance Information Ambiguity". *The Academy of Management Journal* 40.3, pp. 603–625.

Hilger, N. G. (2016). "Why Don't People Trust Experts?" *The Journal of Law and Economics* 59.2, pp. 293–311.

Holzer, H. and D. Neumark (2000). "Assessing Affirmative Action". *Journal of Economic Literature* 38.3, pp. 483–568.

Hyman, R., A. Klarsfeld, E. Ng, and R. Haq (2012). "Introduction: Social regulation of diversity and equality". *European Journal of Industrial Relations* 18.4, pp. 279–292.

Ibanez, M. and G. Riener (2018). "Sorting through Affirmative Action: Three Field Experiments in Colombia". *Journal of Labor Economics* 36.2, pp. 437–478.

Ifcher, J. and H. Zarghamee (2016). "Do gender-variant preferences for competition persist in the absence of performance?" *Economic Inquiry* 54.4, pp. 1918–1930.

Ip, E., A. Leibbrandt, and J. Vecci (2020). "How Do Gender Quotas Affect Workplace Relationships? Complementary Evidence from a Representative Survey and Labor Market Experiments." *Management Science* 66.2, pp. 503–1004.

Iriberri, N. and P. Rey-Biel (2019). "Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-stage Competition in Mathematics". *Economic Journal* 129.620, pp. 1863–1893.

Jensenius, F. R. (2015). "Development from Representation? A Study of Quotas for the Scheduled Castes in India". *American Economic Journal: Applied Economics* 7.3, pp. 196–220.

Karle, H., G. Kirchsteiger, and M. Peitz (2015). "Loss aversion and consumption choice: Theory and experimental evidence". *American Economic Journal: Microeconomics* 7.2, pp. 101–20.

Kato, T. and P. Shu (2016). "Competition and social identity in the workplace: Evidence from a Chinese textile firm". *Journal of Economic Behavior & Organization* 131.Part A, pp. 37–50.

Kerschbamer, R., M. Sutter, and U. Dulleck (2017). "How Social Preferences Shape Incentives in (Experimental) Markets for Credence Goods". *Econ J (London)* 127.600, pp. 393–416.

Kerschbamer, R., D. Neururer, and M. Sutter (2016). "Insurance coverage of customers induces dishonesty of sellers in markets for credence goods". *Proceedings of the National Academy of Sciences* 113.27, pp. 7454–7458.

Kerschbamer, R. and M. Sutter (2017). "The Economics of Credence Goods–A Survey of Recent Lab and Field Experiments". *CESifo Economic Studies* 63.1, pp. 1–23.

Kölle, F. (2017). "Affirmative action, cooperation, and the willingness to work in teams". *Journal of Economic Psychology* 62, pp. 50 –62.

Lane, T. (2016). "Discrimination in the laboratory: A meta-analysis of economics experiments". *European Economic Review* 90, pp. 375 –402.

Leibbrandt, A. and J. A. List (2018). "Do Equal Employment Opportunity Statements Backfire? Evidence From A Natural Field Experiment On Job-Entry Decisions". *NBER Working Paper* 25035.

Leibbrandt, A., L. C. Wang, and C. Foo (2017). "Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women". *Management Science* 64.8, pp. 3501–3516.

Leonard, K., J. List, and U. Gneezy (Sept. 2009). "Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society". *Econometrica* 77, pp. 1637–1664.

Maggian, V. and N. Montinari (2017). "The spillover effects of gender quotas on dishonesty". *Economics Letters* 159.C, pp. 33–36.

Matsa, D. A. and A. R. Miller (2013). "A Female Style in Corporate Leadership? Evidence from Quotas". *American Economic Journal: Applied Economics* 5.3, pp. 136–169.

McGuire, T. G. (2000). "Physician agency". *Handbook of Health Economics*. Ed. by J. C. Anthony and P. N. Joseph. Vol. 1, Part A. Elsevier. Chap. 9, pp. 461–536.

McKinsey & Company (2019). *Women in the workplace*. Report, McKinsey & Company, New York.

Mimra, W., A. Rasch, and C. Waibel (2016). "Price competition and reputation in credence goods markets: Experimental evidence". *Games and Economic Behavior* 100, pp. 337–352.

Mollerstrom, J. (Sept. 2019). "Favoritism and cooperation". *Public Choice*. forthcoming.

Niederle, M., C. Segal, and L. Vesterlund (2013). "How costly is diversity? affirmative action in light of gender differences in competitiveness." *Management Science* 59.1, pp. 1–16.

Niederle, M. and L. Vesterlund (Aug. 2007). "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122.3, pp. 1067–1101.

— (2011). "Gender and Competition". *Annual Review of Economics* 3.1, pp. 601–630.

Pande, R. (2003). "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India". *American Economic Review* 93.4, pp. 1132–1151.

Petters, L. and M. Schroeder (2020). "Negative Side Effects of Affirmative Action: How Quotas Lead to Distortions in Performance Evaluation". *European Economic Review* In Press, Journal Pre-proof, p. 103500.

Rasch, A. and C. Waibel (2018). "What Drives Fraud in a Credence Goods Market? Evidence From a Field Study". *Oxford Bulletin of Economics and Statistics* 80, pp. 605–624.

Raven, J. and C. Raven (2008). "Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics." Competency Motivation Project. Chap. General Introduction and Overview: The Raven Progressive Matrices Tests: Their Theoretical Basis and Measurement Model.

Rotter, J. B. (1966). "Generalized expectancies for internal versus external control of reinforcement". *Psychological Monographs: General and Applied* 80.1, pp. 1–28.

Schildberg-Hörisch, H., C. Trieu, and J. Willrodt (Apr. 2020). "Perceived fairness and consequences of affirmative action policies". *IZA Discussion Paper* 13202.

Schneider, H. S. (2012). "Agency Problems and Reputation in Expert Services: Evidence From Auto Repair". *The Journal of Industrial Economics* 60.3, pp. 406–433.

Schotter, A. and K. Weigelt (May 1992). "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results*". *The Quarterly Journal of Economics* 107.2, pp. 511–539.

Schupp, J. and J.-Y. Gerlitz (2008). "BFI-S: Big Five Inventory-SOEP". *In Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version.* 12.

SGB IX – *German Social Code, Volume 9* (2001). *§ 154 SGB IX Pflicht der Arbeitgeber zur Beschäftigung schwerbehinderter Menschen (Employers' duty of employing people with disabilities)*. https://www.sozialgesetzbuch-sgb.de/sgbix/154.html.

Sowell, T. (2004). *Affirmative Action around the World: An Empirical Study*. Yale UniversityPress.

Sutter, M., D. Glätzle-Rützler, L. Balafoutas, and S. Czermak (2016). "Cancelling out early age gender differences in competition: an analysis of policy interventions". *Experimental Economics* 19.2, pp. 412–432.

Sutter, M. and D. Glätzle-Rützler (2015). "Gender Differences in the Willingness to Compete Emerge Early in Life and Persist". *Management Science* 61.10, pp. 2339–23354.

Sutter, M., C. Zoller, and D. Glätzle-Rützler (2019). "Economic behavior of children and adolescents – A first survey of experimental economics results". *European Economic Review* 111.C, pp. 98–121.

Taylor, C. R. (1995). "The Economics of Breakdowns, Checkups, and Cures". *Journal of Political Economy* 103.1, pp. 53–74.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). "The German Socio-Economic Panel Study (SOEP) - evolution, scope and enhancements". *SOEPpapers on Multidisciplinary Panel Data Research, No. 1*.

Declaration of Authorship

Eidesstattliche Versicherung

Ich, Chi TRIEU, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine Universität Düsseldorf" erstellt worden ist.

Düsseldorf, der 28. September 2020

Chi Trieu