# Essays in Behavioral and Labor Economics

Von der Wirtschaftswissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

- Doctor rerum politicarum -

genehmigte Dissertation

von

Dipl.-Vw. Alexander Straub

geboren am 10.11.1984 in Jaroslawl

2019

Referent: Prof. Dr. Patrick P. Puhani, Leibniz Universität Hannover

Koreferent: Prof. Dr. Stephan L. Thomsen, Leibniz Universität Hannover

Tag der Promotion: 03.09.2019

## Acknowledgements

This dissertation comprises many years of work and taught me various aspects for which I am very thankful to a lot of people without whom this would not have been possible.

First of all, I would like to thank Patrick Puhani for giving me the opportunity to do research under his supervision. He enabled me to try out different things and to do research on the topics which I was interested in. Continuously, he encouraged me to improve my research by taking the time to discuss my progress and providing well-intentioned criticism to polish my papers. He provided an excellent environment with fair working conditions and a friendly and productive atmosphere at the whole institute. He also gave me the opportunity and encouraged me to visit international conferences, where I could discuss my research and motivate myself to further improve it. His solution-oriented manner served me as a blueprint to focus on the important things and to use my time more efficiently.

I also thank Stephan Thomsen, my second supervisor, for his advice on my research. Whenever we discussed my academic work and our joint course in *Applied Econometrics* he gave me all the freedom to develop according to my preferences while constantly providing feedback in a very constructive way. In his friendly and humorous nature, he provided many anecdotes which either supported the argument of my papers or pointed out their limitations, thereby always deepening my understanding of the topic in an enjoyable way.

My experience as a graduate student would be much poorer without my colleagues from the Institute of Labor Economics and the Leibniz Universität Hannover. I always enjoyed to share offices with you, to develop and pursue research ideas, and to have inspiring conversations during coffee breaks. Whenever I needed advice I could benefit from your experience.

I would also like to thank my family for their ongoing support. I could always count on the help of my parents. My wife, Anastasia, was very patient with me and my working hours. Being able to discuss my research with you helped me a lot.

# Abstract

This thesis analyzes the role of some aspects in human behavior and their implications for labor market outcomes using data from laboratory experiments and secondary data. The first chapter outlines the relevance of the topic and further discusses possible policy implications.

The second chapter investigates gender-specific differences in competitive behavior and performance under pressure in non-professional sports. Analyzing data from German ninepin bowling, a mixed-gender sport with direct competition, there is no evidence for gender differences in competing against the opposite sex, once controlling for ability. Gender differences in tight situations also do not play a role.

The third chapter studies the determinants of distributional preferences and their relationship to wages. The first step documents a relationship between distributional preferences revealed in an incentivized experiment and personality traits. In the second step, I provide indicative evidence that inequity aversion and spitefulness decrease earnings in Germany and can partly explain the gender wage gap.

The fourth chapter investigates discrimination based on appearance and religious practice of job applicants using data from a laboratory experiment. There are heterogeneous effects of headscarf on callback rates based on types of occupations and recruiters' characteristics. Better labor market characteristics of applicants can overcompensate discrimination against headscarf.

The fifth chapter analyses how both own appearance and the beauty composition of other candidates influence the chances of being selected for a job interview. Based on a laboratory experiment with randomized CVs, it shows that appearance of other applicants with the same gender has significant incremental effects on top of the existing beauty premium.

**Keywords:** competitiveness, distributional preferences, labor market discrimination

# Zusammenfassung

Diese Dissertation untersucht die Rolle einzelner Aspekte menschlichen Verhaltens und deren Auswirkungen auf Arbeitsmarktergebnisse unter Zuhilfenahme von Experimental- und Sekundärdaten. Das erste Kapitel stellt die Relevanz des Themas heraus und zeigt mögliche politische Implikationen auf.

Im zweiten Kapitel wird untersucht, ob sich Männer und Frauen in Wettbewerbsverhalten und Leistungsfähikeit unter Druck im Freizeitsport voneinander unterscheiden. Unter Verwendung von Daten aus dem Kegelsport, der sich durch gemischtgeschlechtlichen und direkten Wettbewerb auszeichnet, finden sich keine Unterschiede im Wettbewerbsverhalten gegen das andere Geschlecht, sobald individuelle Fähigkeiten berücksichtigt werden. Die Leistungsfähikeit von Männern und Frauen unter Druck unterscheidet sich ebenfalls nicht.

Das dritte Kapitel behandelt die Determinanten von Verteilungspräferenzen und deren Beziehung zu Löhnen. Im ersten Schritt werden in einem anreizkonformen Experiment offenbarte Verteilungspräferenzen in Bezug zu Persönlichkeitsmerkmalen gesetzt. Im zweiten Schritt werden indikative Belege aufgezeigt, dass Ungleichheitsaversion und Boshaftigkeit die Einkommen in Deutschland verringern und teilweise geschlechtsspezifische Lohnunterschiede erklären können.

Das vierte Kapitel untersucht unter Verwendung von Daten aus einem Laborexperiment, ob Aussehen und Religionsausübung von Bewerbern Einfluss auf Einstellungsentscheidungen haben. Es finden sich, je nach Berufsbild und Charakteristiken der Entscheider, heterogene Effekte von Verschleierung auf die Wahrscheinlichkeit, zum Vorstellungsgespräch eingeladen zu werden. Bessere Arbeitsmarktcharakteristiken der Bewerber können Diskriminierung aufgrund des Kopftuches überkompensieren.

Das fünfte Kapitel analysiert, wie sowohl das eigene Erscheinungsbild als auch die Attraktivität von Mitbewerber die eigenen Chancen beeinflussen, zu einem Vorstellungsgespräch eingeladen zu werden. Basierend auf einem Laborexperiment mit randomisierten Lebensläufen zeigt sich, dass das Erscheinungsbild der gleichgeschlechtlichen Mitbewerber zusätzlich zum eigenen Schrönheitsbonus einen inkrementellen Erklärungsbeitrag liefert.

**Schlagworte:** Wettbewerbsverhalten, Verteilungspräferenzen, Arbeitsmarktdiskriminierung

# Contents

# List of Tables

# List of Figures

CHAPTER 1

# Introduction

# 1 Introduction

*Homo economicus* is a key concept in economics and portrays the human being as strictly rational, self-interested, and consistent. The reality of decision making is more complex and not always characterized by rational behavior. Evidence from behavioral economics complements standard economic theory by examining irrational behavior, cognitive biases, and fairness perceptions. Nobel prizes for behavioral economists – Daniel Kahneman and Vernon L. Smith in 2002, Robert J. Shiller in 2013, and Richard Thaler in 2017 – underline the recognition of this discipline within economics.

Understanding when standard economic theory accurately describes the decision making process and when human behavior deviates from *homo economicus* is crucial for effective policy design. In particular, gender gaps in the labor market and discrimination in the hiring process are persisting problems in today's society. For example, Boll and Lagemann (2018) report a still unexplained gender wage gap of 6% in Germany and Weichselbaumer (2016) finds 3 to 4.5 times lower callback rates for applicants wearing a headscarf in a field experiment in Germany. Despite extensive evidence on the existence of these problems, it is still not entirely understood why gender gaps exist and where discrimination arises. Causal evidence is necessary to address these issues and to develop targeted policy actions.

This thesis contributes to the existing literature by analyzing peculiarities of human behavior and their implications for the labor market. First, it aims at shedding light on whether cognitive biases as a form of irrationality alter competitive behavior. Second, I add evidence on the self-interested nature of human beings by analyzing gender differences in other-regarding preferences. The last two chapters are dedicated to consistency and rationality in the hiring process.

Chapter 2 investigates the effect of competing against the opposite sex and gender-specific performance under pressure in a non-professional environment. Using secondary data from around 11,000 German ninepin bowling games comprising over 500,000 observations allows to analyze the gender gaps in competition in a sport where men and women directly compete against each other. To identify the causal effect of playing against the opposite gender on performance, opponent team's female share is used as an instrument. Systematic differences in playing against the opposite gender are not evident once controlling for ability. The non-existing gender gap is precisely estimated due to the large number of observations. Analyzing tight game situations, no evidence for gender-specific differences in performance under pressure can be found.

In chapter 3, I investigate gender differences in other-regarding preferences and their role for wages and subsequently the gender pay gap. First, I show that gender is an important determinant for distributional preferences in an incentivized experiment. In a second step experimental data is coupled with data from SOEP, a representative German panel dataset, to provide indicative evidence that inefficient distributional preferences are associated with lower earnings and can partly explain the gender pay gap.

Chapter 4 exploits the German practice of sending a photo with a CV. It draws on a laboratory experiment to empirically investigate the effect of appearance, ethnicity and religious practice of fictive job applicants on their hiring chances. Student participants are asked to pre-select fictitious candidates for various occupations from a pool of CVs with comparable characteristics but with different photos. The setting allows to identify situations where discrimination occurs, e.g. which job positions and characteristics of the applicants and of the "recruiters" are relevant. Interestingly, there are significant and robust interaction effects between the characteristics of a CV and wearing a headscarf.

Chapter 5 assesses the impact of both own appearance and the beauty composition of other candidates on the chances of being selected for a job interview. To identify the causal effect, photos are randomly included in a pool of CVs and job openings in the laboratory experiment. This allows to create a "wingman" variable based on each applicant's beauty relative to other candidates of the same gender. There are significant incremental effects of the wingman beauty on top of the existing beauty premium. This effect is more pronounced in high skilled occupations and prevails mainly among male recruiters. The findings also suggest that such effects are not driven by intuitive decision processes and predominantly affect decisions at the margin.

Summarizing the findings of this thesis, there is no evidence that cognitive biases as a form of irrationality alter competitive behavior in a non-professional environment without monetary rewards once controlling for ability. "Intrinsic" motivation to win does not seem to differ by gender. If gender gaps in competitiveness arise only in situations with high monetary rewards, this can be seen as an indicator that women value money less. Conversely, assuming human beings to be strictly self-interested may fall short of reality as other-regarding preferences exist and significantly differ by gender. Providing a link between inefficient distributional preferences and lower wages can help to shed light on the reasons behind the still unexplained gender pay gap. Finally, taste based discrimination in the hiring process suggests that consistency and rationality do not always guide human behavior. Decision makers in the experiment predominately discriminate against headscarf when applicants have inferior characteristics, whereas good characteristics revert the biased perception possibly through implicit affirmative action. Policy interventions such as education or apprenticeship programs targeting minority groups might effectively prevent discrimination in the hiring process. Discrimination based on appearance is amplified in situations where decisions are at the margins and the gender ratio is unbalanced. A large pool of candidates with the same gender can reduce discrimination based on appearance, e.g. by encouraging balanced gender ratios across occupations.

CHAPTER 2

# Mind the absent gap: Gender-specific competitiveness in non-professional sports

# 2 Mind the absent gap: Gender-specific competitiveness in non-professional sports

*Co-authored with Anna Katharina Pikos.*

## 2.1 Introduction

*Man differs from woman in size, bodily strength, hairyness, &c., as well as in mind, in the same manner as do the two sexes of many mammals.*

– Charles Darwin, *The Descent of Man, 1871, pp.13-14*

Do men compete in the same way as women? Men earn higher wages, get more promotions, and hold the majority of leadership positions. This holds also when accounting for standard economic variables. For example, the wage gap persists for women with full-time employment history, without children, and without family plans (e.g Manning and Swaffield, 2008). Similarly, the promotion gap is not fully explained by worker performance and firm characteristics (e.g Blau and DeVaro, 2007). This does not necessarily prove discrimination as there might be differences in unobservable characteristics.

Personality differences received rising scholarly attention as a potential explanation for the persisting gender gaps in the labour market. The experimental literature finds that men are more prone to select into competitive environments (Niederle and Vesterlund, 2007; Dohmen and Falk, 2011), are less risk-averse (Croson and Gneezy, 2009), and perform better when competing against women (Antonovics et al., 2009, significant at the 10% level). Women seem to increase their performance less under mixed-gender competition (Gneezy et al., 2003). The sports economics literature focuses on differences in competitive behavior of high performers and identifies risk behavior, environment, and stakes as important factors. Men take more risks when risky behavior might pay off (Böheim et al., 2016) and when it does not (Gerdes and Gränsmark, 2010), women perform better in a female environment (Booth and Yamamura, 2018), and men choke under pressure (Cohen-Zada et al., 2017). Incentives in professional sports are predominantly monetary. Insights about the importance of non-monetary rewards in gender-specific competitiveness are still missing but would help to better understand "intrinsic" motivation.

We aim at filling this gap by analyzing gender differences in a non-professional environment with non-monetary rewards. Our data comes from German ninepin bowling, a mixed-gender sport where individual players compete against one opponent to obtain points for their team. This data has three advantages. First, men and women compete against each other in the same environment. Second, unobserved group dynamics exert less influence in a one-against-one competition. Third, our panel data allows to control for past performance as a proxy for ability. This enables us to isolate the effect of competing against the opposite gender on performance. Our dataset consists of over 500,000 observations from more than 11,000 games. About 15% of the games are from mixed-gender leagues.

We answer two questions: first, do men and women compete differently against the opposite gender and second, are there systematic gender differences in performance under pressure? Descriptive statistics document that men perform slightly better and hence are more likely to win against women. Any differences in playing against the opposite gender are fully explained by ability or game characteristics. We confirm the OLS findings with fixed effects and IV. To rule out the possibility of non-random matching completely we instrument the probability of playing against the opposite gender with the gender composition of the opponent team. To address gender differences in performance under pressure, we analyze tight game situations. We do not find any evidence for a gender performance gap under pressure.

The remainder of the paper is organized as follows. Section 2.2 reviews the relevant literature. The data is presented in section 2.3, the estimation strategy in section 2.4. Section 2.5 contains the results, the last section concludes.

## 2.2 Literature

Competition is important for a variety of labor market outcomes such as wages and promotion. Card et al. (2016) find sorting and bargaining effects for premium pay among Portuguese workers. Women are less likely to sort into companies paying higher premiums and they only receive 90% of their male colleagues' premium. There is evidence that part of the gender gaps might be related to different performance in negotiations. When applying for new jobs, women are more inclined to accept lower wages and men are more likely to negotiate wages (Leibbrandt and List, 2015). The experimental literature supports gender differences in negotiations; e.g. women negotiate equally well as employers but make lower initial offers as employees (Dittrich et al., 2014).

Laboratory and field experiments shed light on how men and women behave in competitive environments. Gneezy and Rustichini (2004) for example show that boys and girls in Israel react differently to competition while running. Boys run faster when they face a direct opponent compared to running alone. This is not the case for girls whose performance does not increase under competition. According to Dreber et al. (2011), culture and task play a role, as the effect is less pronounced in Sweden and in rather gender neutral activities, such as skipping a rope. Laboratory experiments confirm that women shy away from competition if possible and do not increase their performance if forced to compete (e.g. Gneezy et al., 2003; Niederle and Vesterlund, 2007). The environment's gender composition is important but empirical evidence is mixed. Booth and Nolen (2012) document that a purely female environment may enhance women's competitive behavior, while Lee et al. (2014) find a non-significant reduction in competitiveness in single-sex environments (see also Niederle and Vesterlund, 2011, for an extensive literature review).

Numerous studies from sports economics confirm a gender gap in competitiveness. Men perform better in mixed-sex Japanese speedboat races, while the opposite is true for women (Booth and Yamamura, 2018). As speedboat races are group races, there might be spillover effects impacting individual behavior that might go in both directions, e.g. racing faster because somebody else is fast or slowing down because somebody cuts the lane. Men are less risk averse when risky behavior might pay off in professional basketball (Böheim et al., 2016) but they perform worse when stakes are very high in tennis (Cohen-Zada et al., 2017). Czibor et al. (2019) show that women are more risk averse in traditionally male environments, even as they gain experience. They use data from an online card game where men and women compete directly against each other. This eliminates group spillover effects and league differences. Different leagues have different characteristics (e.g. wage or quality of medical treatment in case of injury) which might impact players' behavior, e.g. National Basketball Association versus Women National Basketball Association, and Grand Slam tournaments. Gerdes and Gränsmark (2010) take data from mixed-gender chess games and find that men not only are more risk-loving but also choose more aggressive opening strategies against women even if this reduces their probability to win the game. According to de Sousa et al. (2015) and Backus et al. (2016), women play worse against men in chess competitions.

We lack insights regarding the everyday work life of the general population. Professional sports are very selective and representative for top performers only, e.g. managers. Trophy money, advertising contracts, and prestige are at stake. To address competitiveness beyond the upper end of the ability distribution, non-professional sports are insightful. Analyzing situations when winning is not tied to monetary rewards enables a better understanding of gender differences underlying the “pure” motivation to win. Differences in competitive behavior could be more or less pronounced in non-professional sports.[1]

## 2.3 Data and Descriptives

The data comes from ninepin bowling, a non-professional sport played in many European countries. Even world-class players do not get paid.[2] German ninepin bowling leagues (“Kegeln”) on the national and federal state level are gender-separated. Lower leagues on county level can be mixed-gender to make it easier for small clubs to put together a team. Because there are no financial incentives, we refer to this sport as a low stake environment. Any differences in competitiveness would mostly be intrinsic, especially in the less selective lower leagues we use for this analysis. Reaching a higher league is not associated with financial incentives.

### 2.3.1 Data

This subsections outlines the rules for ninepin bowling leagues regulated by the German ninepin bowling association (“Deutscher Keglerbund Classic e.V.”). Two teams of four or six players bowl against each other. Each player has a direct opponent against whom she bowls in four sets. Each player has 120 throws, i.e. 30 throws per set. For the first fifteen throws of each set the pins reset after each bowl. The sum of knocked down pins is recorded as *V-score*. For the last 15 throws in each set the pins reset only after the player knocked down all the 9 pins, i.e. if a player fails to knock down all of them, she uses up a next throw to fell the remaining pins. We refer to the sum of the cleared pins from the last 15 throws of a set as *A-score*. After the player and his opponent finished their 30 throws *V-score* and *A-score* are added up to the *score*. The

[1]On the one hand, men choke under pressure (Cohen-Zada et al., 2017) and the highest-performing women are more prone to select into competition (Gneezy and Rustichini, 2004). Based on these results differences in non-professional sports could be more pronounced. On the other hand, women perform better in low stake environments (Ors et al., 2013). This could result in a lower gender gap.

[2]There might be sponsoring contracts for European and World championships, but this will hardly affect any of our results because we analyze leagues which are at least 6 levels lower.

player with the highest *score* receives a *point* for winning the set.[3] Thereupon the two players switch the lanes and the next set starts. After the four sets are completed, the player with the highest amount of *points* receives a *team point*.[4] The alleys consist of four lanes where two home and two guest players bowl simultaneously and form a *pairing*. The players switch lanes after each set, so that every player bowls on every lane. When the game is finished, the team with the highest sum of *score* receives two additional *team points*. The team with the highest sum of *team points* wins the game. The losing team does rarely benefit from the earned *team points*.[5]

Our data comes from leagues in the Northern part of the federal state of Baden-Württemberg in Southern Germany. The local bowling association "Württembergischer Kegler- und Bowling-Verband e.V. (WKBV)" publishes game records online. Records contain information on league, location, date, start and end of the game. Player information is limited to name and player number. Performance measures include *points*, *V-score*, *A-score*, and *F-score* (mistakes, i.e. number of bowls not hitting any pin). We use game records from the seasons 2014/15 to 2017/18.

We exclude players playing fewer than five games, games with a predetermined winner[6] and players without opponents. Gender is coded according to Wikipedia lists for male and female given names.[7] We use record sheet information symmetrically from the home player's and the guest player's point of view, i.e. in mixed-gender competition, we have women vs. man and man vs. women. This leaves us with around 75,000 observations from about 2,000 games in mixed-gender leagues. The full sample with gender-separated leagues counts around 500,000 observations from more than 11,000 games.

### 2.3.2 Descriptives

Men bowl significantly better than women in mixed-gender leagues although the differences are not large (see appendix figure A2.1). Table 2.1 shows the raw gender differences in outcomes in mixed-gender games (see appendix table A2.1 for full sample descriptives). On average, men

[3] If both players have the same *score* they receive 0.5 set *points*.

[4] If both players obtained two set points, the one with the highest sum of knocked down pins (total *score* over four sets) gets the team point. If both knocked down the same amount of pins and both have two set points, each of them receives 0.5 team points.

[5] Team points are only important in the ranking if ranking points are the same for two teams. The winning team earns 2:0 ranking points, the losing team 0:2, and 1:1 for ties.

[6] If a team lacks two or more of the scheduled players, there is no possibility to win the game.

[7] About 250 of 3,500 players could not be matched because their names were either not listed on Wikipedia or for unisex first names such as e.g. Robin and Gabriele. We infer the gender for most of those players because they participated at least once in a strictly gender separated tournament or league. For the remaining players, we looked up the gender on the web-page of their club.

bowl 0.5 pins more than women; 0.3 pins in *V-score* and 0.2 pins in *A-score*. Men win their set slightly more often and make somewhat more mistakes but the latter difference is very small and significant only at the 10% level. Table 2.2 gives an overview over the number of observations for own and opponent's gender. There are 75,056 observations. Roughly 48% of the encounters are mixed-gender encounters. Male encounters make up around 31%, female encounters around 21%.[8]

Table 2.1: Outcomes by gender

| | men | women | difference | p-value |
|---|---|---|---|---|
| score | 113.833 | 113.319 | 0.514 | 0.000 |
| points | 0.507 | 0.492 | 0.015 | 0.000 |
| V-score | 80.776 | 80.470 | 0.306 | 0.000 |
| A-score | 33.057 | 32.849 | 0.208 | 0.004 |
| F-score | 3.716 | 3.691 | 0.025 | 0.182 |
| Observations | 41211 | 33783 | | |
| Distinct players | 1016 | 612 | | |

*Notes:* Results for t-tests on differences of game outcomes by gender. See page 90 for a detailed discussion of the variables. Data source: WKBV.

Table 2.2: Numbers of observations for own and opponent's gender

| | opponent's gender | | | | | |
|---|---|---|---|---|---|---|
| own gender | male | | female | | Total | |
| | No. | % | No. | % | No. | % |
| male | 23,314 | 31.1 | 17,906 | 23.9 | 41,220 | 55.0 |
| female | 17,906 | 23.9 | 15,882 | 21.2 | 33,788 | 45.0 |
| Total | 41,220 | 55.0 | 33,788 | 45.0 | 75,008 | 100.0 |

*Notes:* Cross tabulation of gender and opponent's gender in mixed sex leagues. Data source: WKBV.

## 2.4 Empirical Strategy

This section describes our OLS, fixed effects, and IV models. The outcomes are performance measures $y_{ijk}$ of player $i$ against opponent $j$ in environment $k$. The main explanatory variables are

[8]The female share in ninepin bowling (45% in mixed-gender leagues) is more balanced than in other mixed-gender environments analyzed in the literature. Around 8% of the card players in Czibor et al. (2019) are women. In chess, women account for roughly 10% of the players (Gerdes and Gränsmark, 2010; de Sousa et al., 2015; Backus et al., 2016).

gender, playing against the opposite gender, and the interaction term ($female_i$, $opp\_gender_{ij}$, and $female_i \cdot opp\_gender_{ij}$). $Z'_k$ is a vector of "environmental" characteristics $k$ containing dummy variables for *pairing*, *set*, and playing at *home*. $ability'_{ij}$ is a vector of player $i$'s, opponent $j$'s and teams' ability measures.[9] $\varepsilon_{ijk}$ is the error term clustered at players' level.

$$\begin{aligned} y_{ijk} =& \beta_0 + \beta_1 \cdot female_i + \beta_2 \cdot opp\_gender_{ij} + \beta_3 \cdot female_i \cdot opp\_gender_{ij} \\ &+ Z'_k \gamma + Ability'_{ij} \delta + \varepsilon_{ijk} \end{aligned} \tag{2.1}$$

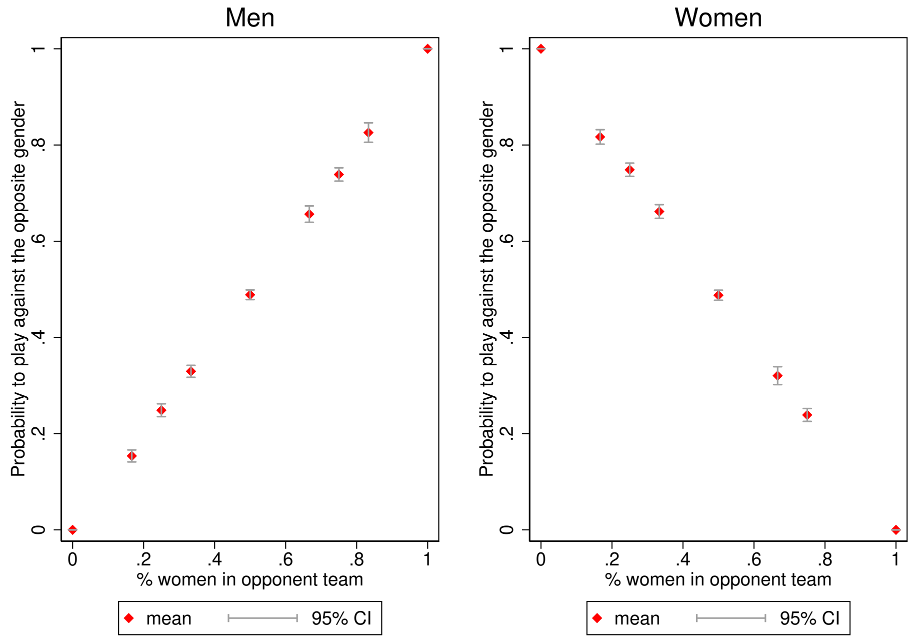
To account for systematic individual and lane differences, we estimate fixed effects models. We validate our results with player fixed effects in case our individual ability measures do not fully capture idiosyncratic differences, e.g. gender-specific returns to experience or panel attrition. There are two reasons to consider lane fixed effects. First, lane quality can vary substantially even within locations. Second, there are lane-specific differences in physical proximity to the other players. When playing on the inner lanes, the current score of the other players might be easier to assess, which could influence own performance.

OLS may yield biased estimates if individuals self-select into playing against opponents of a certain gender, i.e. if they have a preference for competing against men or women. To address the potential endogeneity of $\beta_2$ and $\beta_3$, we use the opponent team's female share as an instrument for playing against the opposite gender. Figure 2.1 displays the raw first stages for rounded values of opponent female share by own gender and confirms the relevance of the instrument. As the share of women in the opposite team increases, the probability to play against the opposite gender increases for men and decreases for women. All first stages are highly significant, the coefficient is virtually 1 in absolute terms, and F-statistics are large (5600 to 9200).

For the exclusion restriction to hold, the opponent team's female share must not have any direct effect on performance. The main concern is that players might have higher confidence when perceiving a predominantly female team as weaker. If indeed the team is weaker, this does not bias the results as long as we control for team and opponent team ability. If the team is not weaker but is only regarded as such, then players have a biased perception of the ability distribution of their opponents by gender. In this case, opponent female share could have a direct effect

[9] We mostly use past performance as a proxy for ability. Alternative measures, such as individual fixed effects and different ability measures based on past and future performance, are used in robustness checks, see table A2.2.

Figure 2.1: Probability to play against the opposite gender for men and women depending on opponent team composition




Rounded values of opponent female share. Excluding games with substitutions. Data source: WKBV.

on performance through false overconfidence (or underconfidence). Given that performance measures on player and team ability are easily available to all players,[10] there should not be any strong biases in perceived opponent ability. Even if players had gender-biased beliefs about performance, they would update their beliefs as soon as they gain experience in playing against both genders in their league. Hence, if such a direct effect existed, it is negligible.

## 2.5 Results

Men do not perform systematically better against women and vice versa (table 2.3). Not accounting for ability, men are 1.5 percentage points more likely to win the set against women.[11] This difference is significant at the 5% level and arises from gender differences in ability. There are no significant gender differences for *score* and *F-score*. The models explain between 0.4% and 2% of the variation in the outcomes. Including ability controls in the right hand part of the

[10]Team and individual performance rankings are distributed online after each matchday. The rankings contain the weekly top 10 separated by home and guest games and cumulative home and guest performance of all players in the league.

[11]Due to the symmetry of our dataset, it is not possible to disentangle whether men perform better against women, whether women perform worse against men, or whether a combination of both is true.

table turns the opponent gender difference for points insignificant. The models explain larger shares of the variation in the outcome (between 18% for *points* and 36% for *score*). The results are robust to different ability measures (see appendix table A2.2).

Table 2.3: OLS estimates for score and points and F-score

| | without ability | | | with ability | | |
|---|---|---|---|---|---|---|
| | score | points | F-score | score | points | F-score |
| female | -0.145 | 0.000 | -0.055 | -0.081 | -0.000 | -0.051 |
| | (0.683) | (0.013) | (0.095) | (0.210) | (0.006) | (0.036) |
| opp. gender | 0.165 | $0.015^{**}$ | -0.025 | -0.096 | 0.006 | 0.002 |
| | (0.262) | (0.006) | (0.036) | (0.157) | (0.005) | (0.025) |
| female × opp. gender | -0.490 | $-0.031^{***}$ | 0.027 | 0.065 | -0.012 | -0.056 |
| | (0.409) | (0.010) | (0.057) | (0.244) | (0.008) | (0.037) |
| $2^{nd}$ set | $0.786^{***}$ | | $-0.102^{***}$ | $0.706^{***}$ | | $-0.095^{***}$ |
| | (0.130) | | (0.021) | (0.136) | | (0.022) |
| $3^{rd}$ set | $0.676^{***}$ | | $-0.103^{***}$ | $0.664^{***}$ | | $-0.092^{***}$ |
| | (0.135) | | (0.022) | (0.142) | | (0.023) |
| $4^{th}$ set | $0.925^{***}$ | | $-0.143^{***}$ | $0.911^{***}$ | | $-0.151^{***}$ |
| | (0.150) | | (0.024) | (0.158) | | (0.025) |
| $2^{nd}$ pairing | $1.037^{***}$ | | $-0.167^{***}$ | -0.157 | | -0.007 |
| | (0.354) | | (0.048) | (0.146) | | (0.025) |
| $3^{rd}$ pairing | $6.155^{***}$ | | $-0.907^{***}$ | -0.141 | | -0.051 |
| | (0.582) | | (0.079) | (0.229) | | (0.038) |
| home | $1.671^{***}$ | $0.058^{***}$ | $-0.182^{***}$ | $1.562^{***}$ | $0.057^{***}$ | $-0.167^{***}$ |
| | (0.174) | (0.004) | (0.024) | (0.154) | (0.004) | (0.021) |
| past ability | | | | $0.832^{***}$ | -0.000 | $-0.110^{***}$ |
| | | | | (0.015) | (0.000) | (0.003) |
| difference ability | | | | -0.004 | $0.015^{***}$ | -0.001 |
| | | | | (0.009) | (0.000) | (0.001) |
| team ability | | | | 0.018 | $0.001^{***}$ | -0.004 |
| | | | | (0.016) | (0.000) | (0.003) |
| opponent team ability | | | | $-0.047^{***}$ | $-0.001^{***}$ | $0.005^{**}$ |
| | | | | (0.014) | (0.000) | (0.002) |
| constant | $110.899^{***}$ | $0.471^{***}$ | $4.121^{***}$ | $22.663^{***}$ | $0.473^{***}$ | $16.220^{***}$ |
| | (0.504) | (0.008) | (0.073) | (1.400) | (0.038) | (0.252) |
| Observations | 75008 | 75008 | 74994 | 64690 | 64690 | 64679 |
| Distinct players | 1628 | 1628 | 1628 | 1471 | 1471 | 1471 |
| Adj. $R^2$ | 0.020 | 0.004 | 0.016 | 0.364 | 0.182 | 0.275 |

*Notes:* This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues. The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. *Female* and *opp. gender* are dummy variables if the player is female or plays against the opposite gender respectively. *Set* and *pairing* are a set of dummy variables that indicate the difference in *score* and *F-score* compared to $1^{st}$ set and $1^{st}$ pairing; these are omitted for *points* due to the symmetry of the data. *Past ability* is the average *score* of the player per lane if more than 8 lanes are observable from past data. *Difference* is the difference between *past ability* of the player and her opponent. *Team ability* and *opponent team ability* are measures for team's quality, they are calculated by the average of *past ability* of other players in the team. Robust standard errors clustered at the level of the player are in parentheses. $^{***}$, $^{**}$ and $^{*}$ denote significance at the 1%, 5% and 10% level, respectively. Data source: WKBV.

To account for idiosyncratic differences across players and lanes, we run fixed effects regressions (see appendix table A2.3). With individual fixed effects, there is only one significant gender difference: women score 0.3 fewer pins when playing against men (significant at the 10% level). There is no difference for *points* and *F-score*. All gender coefficients are insignificant with lane fixed effects. The same is true when combining individual and lane fixed effects.

Instrumenting playing against the opposite gender with opponent team's female share confirms our results. Table 2.4 displays the second stages; first stage t-statistics can be found at the bottom of the table. Men do not compete differently against women in any outcome but women make 0.14 fewer mistakes when playing against men compared to playing against women. This difference is significant at the 5% level. Hence, there is no evidence that our OLS results are driven by selection into competing against the opposite gender.

Table 2.4: Second stage IV estimates by gender

| | score | | points | | F-score | |
|---|---|---|---|---|---|---|
| | women | men | women | men | women | men |
| opp. gender | 0.641 | -0.382 | -0.016 | 0.008 | -0.137** | 0.071 |
| | (0.418) | (0.314) | (0.011) | (0.010) | (0.061) | (0.052) |
| home | 2.024*** | 1.180*** | 0.063*** | 0.052*** | -0.209*** | -0.132*** |
| | (0.235) | (0.198) | (0.006) | (0.005) | (0.033) | (0.027) |
| past ability | 0.819*** | 0.838*** | 0.000 | -0.000 | -0.110*** | -0.110*** |
| | (0.020) | (0.023) | (0.001) | (0.001) | (0.004) | (0.004) |
| difference ability | 0.012 | -0.014 | 0.016*** | 0.015*** | -0.002 | 0.000 |
| | (0.013) | (0.012) | (0.000) | (0.000) | (0.002) | (0.002) |
| team ability | 0.010 | 0.025 | 0.001 | 0.001** | -0.001 | -0.008** |
| | (0.024) | (0.022) | (0.001) | (0.001) | (0.004) | (0.003) |
| opponent team ability | -0.031 | -0.056*** | -0.001** | -0.001** | 0.002 | 0.006** |
| | (0.020) | (0.019) | (0.001) | (0.001) | (0.003) | (0.003) |
| constant | 22.556*** | 22.342*** | 0.490*** | 0.466*** | 16.131*** | 16.279*** |
| | (2.049) | (1.879) | (0.062) | (0.047) | (0.404) | (0.323) |
| Observations | 29130 | 35560 | 29130 | 35560 | 29126 | 35553 |
| Distinct players | 569 | 902 | 569 | 902 | 569 | 902 |
| Adj. $R^2$ | 0.372 | 0.357 | 0.190 | 0.175 | 0.278 | 0.272 |
| First stage t-statistic | -75.295 | 96.065 | -74.820 | 95.898 | -75.295 | 96.065 |

*Notes:* This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues. The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. Models control for *set* and *pairing* (except for points as dependent variable), and *home*. *Past ability* is the average *score* of the player per lane if more than 8 lanes are observable from past data. *Difference* is the difference between *past ability* of the player and her opponent. *Team ability* and *opponent team ability* are measures for team's quality, they are calculated by the average of *past ability* of other players in the team. Robust standard errors clustered at the level of the player are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively. Data source: WKBV.

Reaction to opposite gender might differ over sets. The first set is important to get into the game and to make a good start, while the last set is the last opportunity to gain a set point. Gender-specific differences in physical and mental endurance could cancel each other out. To detect such differences we estimate equation (2.1) separately for each set in the full sample of over 500,000 observations (see appendix table A2.1 for descriptive statistics). Figure 2.2 plots the reaction to opposite gender for men and women over the four sets (x-axis). There are no systematic gender differences in any outcome in any set. Effect sizes are very close to zero and confidence intervals almost always include zero.

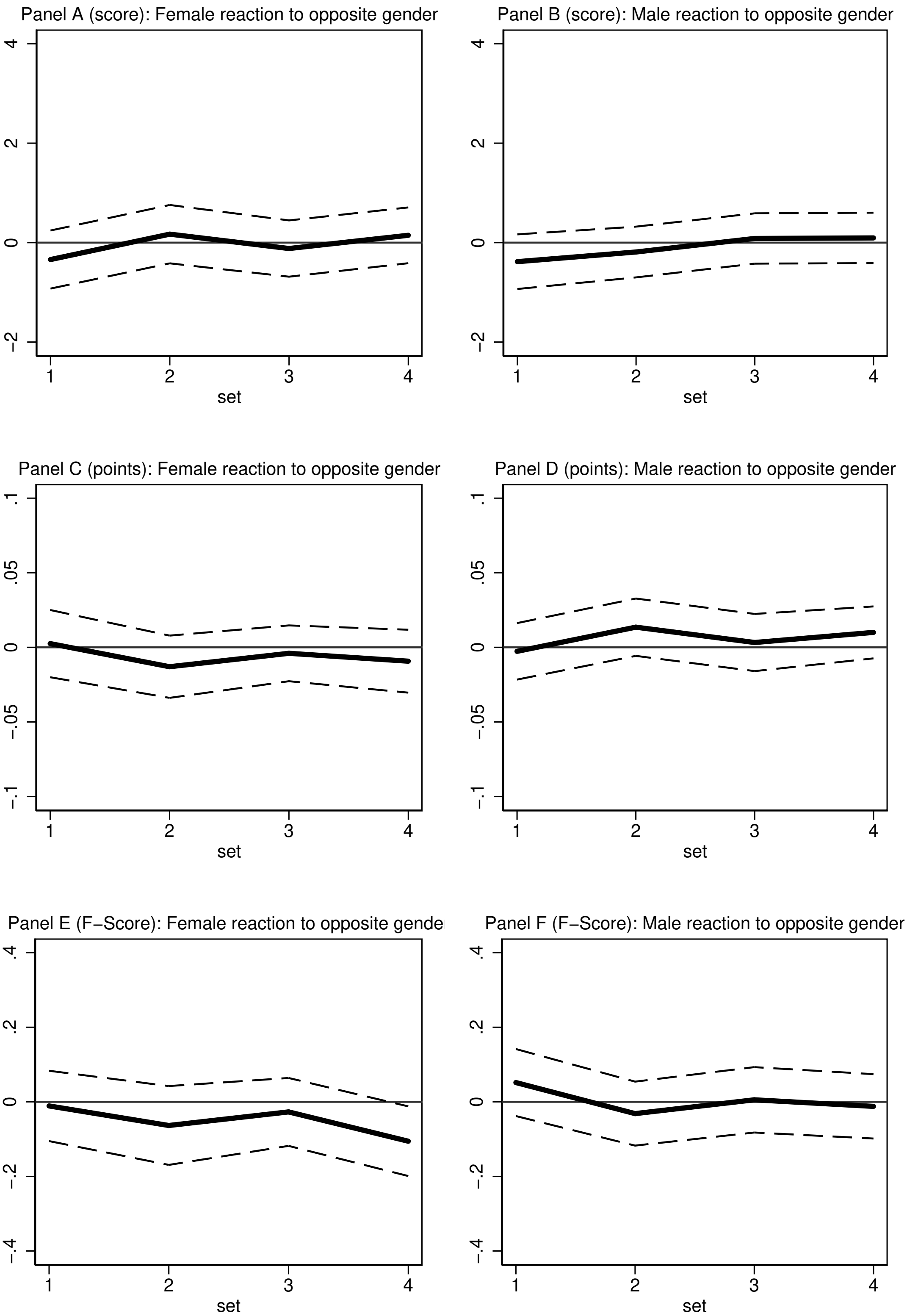
Our results could hide gender differences in performance under pressure. In some situations performance matters more than in others, e.g. turning around a tight game is more important than winning an already decided game. The overall estimate could be zero because one gender might perform better in tight situations but worse when pressure is low. We estimate equation (2.1) in subsamples to consider three different game situations. Figure 2.3 shows the estimated coefficients for the female dummy and their standard errors in three situations of the game. The x-axes are the differences between own and opponent's performance on individual or team level; negative (positive) values indicate that the player/team is lagging behind (leading).[12] First, we analyze performance in the second part of the set (*A-score*, panel A) and the probability to win the set (*points*, panel B) conditional on the difference in the first part of the set (*V-score*). Second, we depict *score* (panel C) and *points* (panel D) depending on the *points* difference from previous sets. Finally, we show *score* (panel E) and *points* (panel F) conditional on the difference in team *points* from previous players.[13]

There is no evidence for any systematic gender differences in playing under pressure. The performance of women lagging behind by up to 10 pins does not differ from their male counterparts (panels A and B). Similarly, we do not find systematic gender differences when lagging behind or leading individually (panels C and D) or as a team (panels E and F). Interestingly, there is no relationship between gender coefficient and the intensity of pressure, which varies along the x-axes, i.e. pressure is higher for smaller differences. For example, in panels C and D when the absolute points difference is three, turning around the game is impossible, while players should feel the highest pressure for absolute differences of one or zero. Overall, there is a reaction to intensity as can be seen by a significant effect of lagging behind or being ahead on subsequent performance (appendix figure A2.2) but this reaction is not gender-specific.

[12] Note that in the analyses underlying figure 2.3, we do not limit the sample to mixed-gender leagues. As we do not find any (causal) effect (see the following), we are confident that this is not biasing the gender dummy we are interested in. The advantage of considering all leagues is that we have larger samples as we run separate regressions for every value on the x-axis of the graphs, e.g. the coefficients displayed in panel A stem from 21 different regressions.

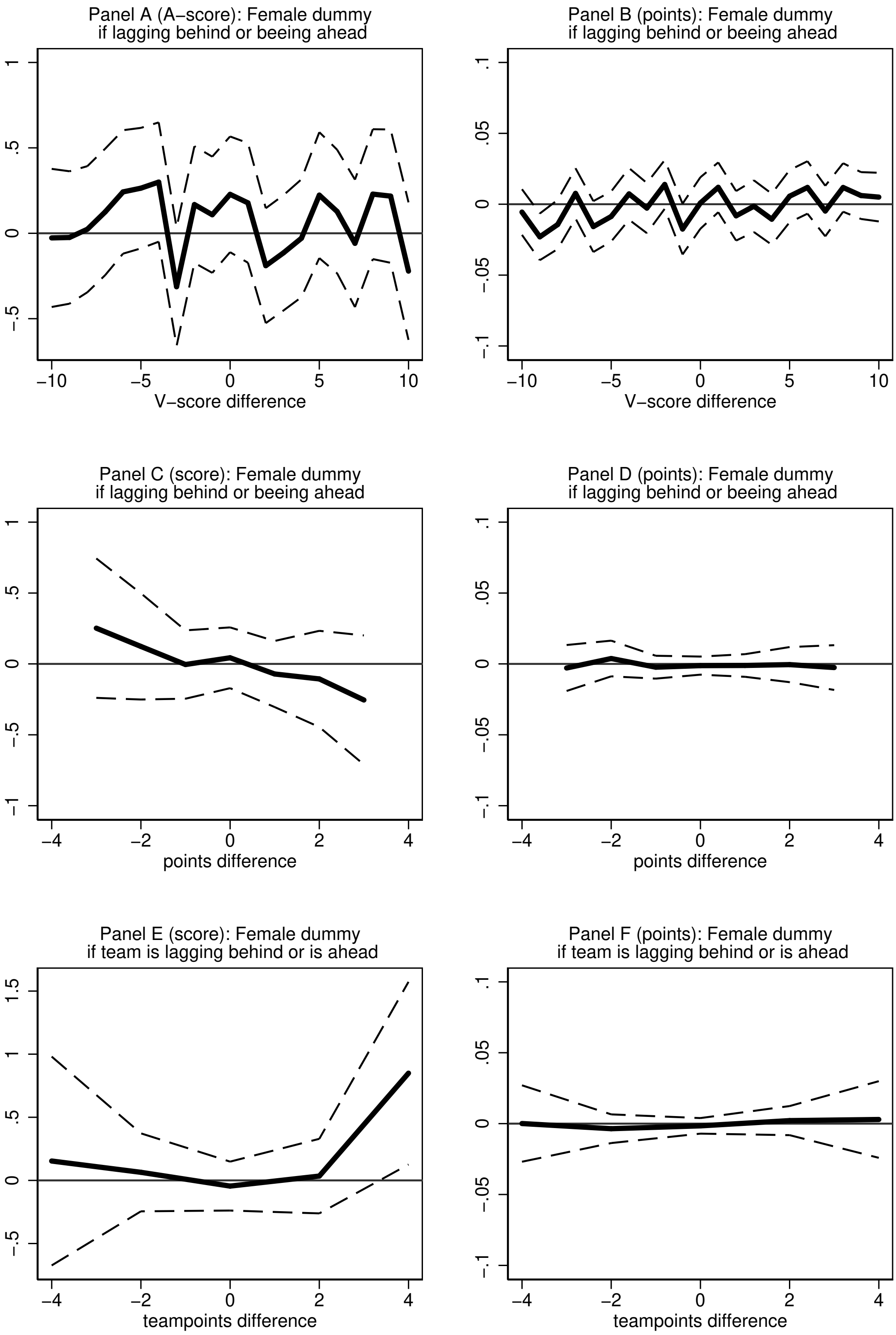
[13] We run regressions only on even values of team point difference since the extremely low number of ties leads to a low number of observations for differences of 1 and 3 team points.

Figure 2.2: Male and female reaction to opposite gender




Dotted lines: 95% confidence intervals. Data sources: WKBV.

Figure 2.3: Gender performance differences in tight situations




Dashed lines: 95% confidence intervals. Data sources: WKBV.

## 2.6 Conclusion

Using a rich panel dataset with direct competition, we analyze competitiveness in non-professional sports and do not find evidence for gender-specific differences. In our setting, participation is less selective, stakes are low, and motivation to win is mainly intrinsic. Our dataset is large ($n > 500,000$) and gives us precisely estimated coefficients with small standard errors. Men play better than women on average but their performance does not depend on their opponent's gender. This is robust across sets and hence not driven by differences in physical and mental endurance. Both genders react similarly to pressure.

There are three possible explanations why our results differ from the literature from professional sports: ability measurement, stakes, and environment. First, the significantly better male competitiveness found in professional sports could stem from measurement error in ability. Our panel dataset allows to construct very detailed ability controls based on past performance. Ignoring gender differences in physical capacities would lead to omitted variable bias and might explain gender differences found in professional sports. As data availability increases, existing studies could be replicated using elaborated proxies for ability, e.g. running distance and speed in tennis and basketball.

Second, the discrepancy in results could arise from stakes. The previous literature exclusively considers high stake environments in which performance is tied to monetary rewards. In situations where mainly pure intrinsic motivation is decisive, men and women compete equally well. If monetary rewards induce men to perform better in mixed-gender settings, this finding could be one part behind the still unexplained gender pay and promotion gap. It also relates to the gender differences in occupational choices, i.e. women self-select into professions with lower expected returns (e.g. nurse, child care worker).

Third, previous works mostly study competitiveness in dynamic and gender-separated environments. In dynamic environments players react to their opponents' strategy and can put pressure on each other more easily, e.g. smashing in tennis. Analyzing individual behavior is more complex as it always depends on opponent behavior. Ninepin bowling is a more static environment in the sense that there are fewer opportunities to put pressure on opponents. This makes it easier to measure individual performance. Gender-separated environments complicate analyses because men and women, although doing the same sports, play in different leagues.

Different leagues come with different characteristics which might impact behavior, e.g. wage or quality of medical treatment in case of injury. Previous studies using data from mixed-gender environments have much lower female representation, e.g. around 8% in an online card game and up to 15% in chess. Ninepin bowling is a more gender balanced environment, where 25% to 45% are women. Children learn the sport in mixed-gender groups, and usually continue to train together as adults. Hence, men and women know from an early age on that they can perform equally well. This reduces false overconfidence from men and excessive underconfidence from women.

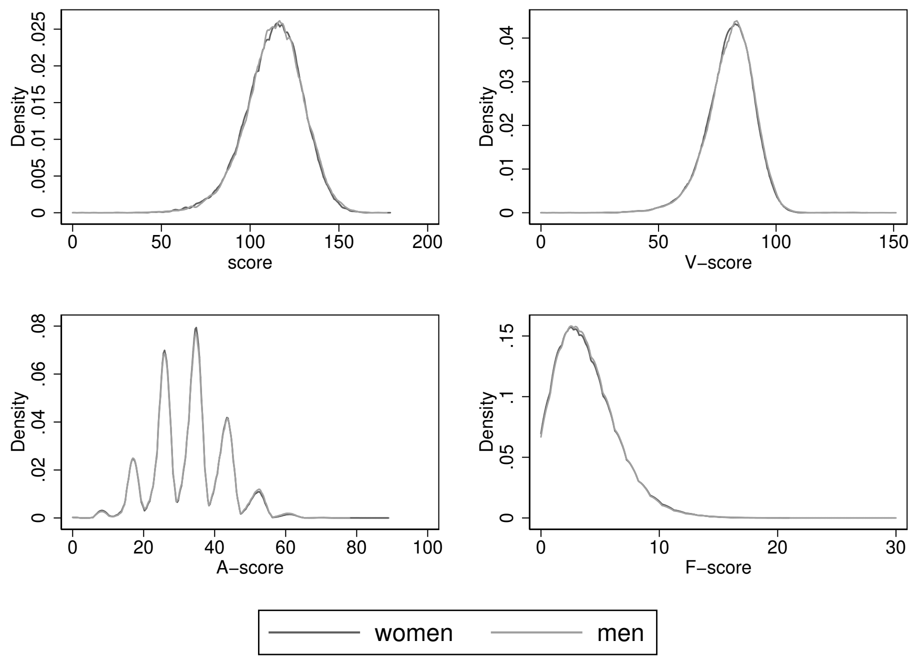
The difference between our result and significant gender gaps in competitiveness from the literature might stem from considering data from different points of the ability distribution. Our data represents a broader population with a more heterogeneous socioeconomic background, while students or professional athletes are a more selective group. The gender gap in willingness to compete might be an upper end phenomenon since there is no gender gap among children from low socioeconomic status (Almås et al., 2015) and among low ability children (Buser et al., 2017).

To conclude we do not find any gender gap in competitiveness in an environment where ability measures is easily observable, monetary incentives do not play a role, and men and women train together. This relates to Exley et al. (forthcoming) who document that women know their value and only enter negotiation when it pays off. In our setting, women assess their own performance correctly because they train in mixed-gender groups from the beginning. If the lack of appropriate information induces a difference in competitiveness between men and women, widely available performance measures might help. If the gender gap arises only in the presence of monetary rewards, this could reflect different preferences, e.g. women might value money less.

## 2.7 Appendix

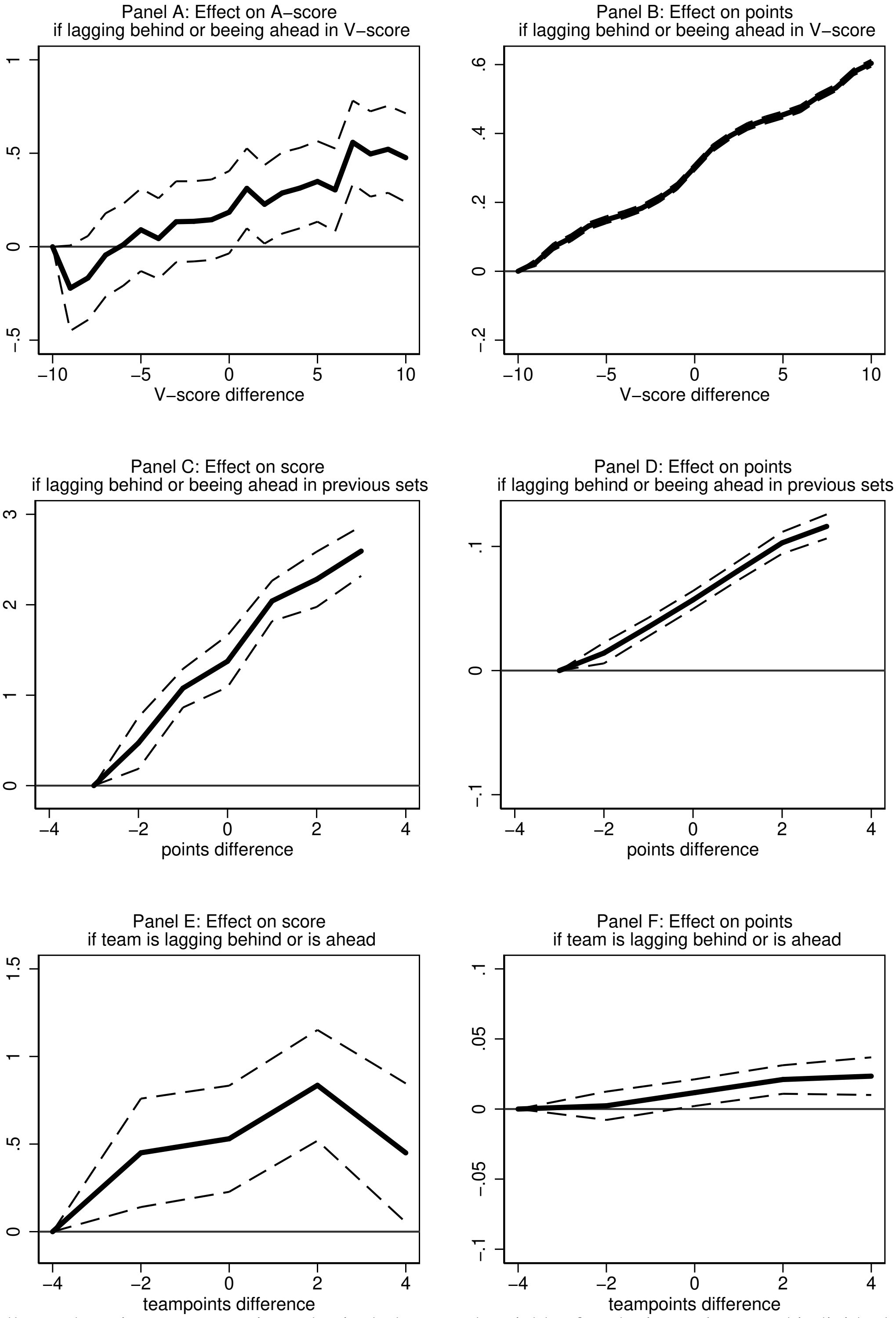
# Appendix Figures

Figure A2.1: Epanechnikov kernel density estimates by gender




Data source: WKBV. Kernel: epanechikov.

Figure A2.2: Performance in tight situations




*Notes:* Full sample estimates, regressions also include control variables for playing at *home* and individual and team ability for player and opponent. Panel A, C, and E also include control variables for *pairing* and *set*. Data sources: WKBV.

# Appendix Tables

Table A2.1: Descriptive statistics (full sample)

| | female | | | | male | | | |
|---|---|---|---|---|---|---|---|---|
| | mean | sd | min | max | mean | sd | min | max |
| score | 121.201 | 14.817 | 0.0 | 184.0 | 126.734 | 14.754 | 0.0 | 195.0 |
| V-score | 84.422 | 8.745 | 0.0 | 114.0 | 87.050 | 8.480 | 0.0 | 146.0 |
| A-score | 36.779 | 9.703 | 0.0 | 89.0 | 39.684 | 10.109 | 0.0 | 91.0 |
| F-score | 2.613 | 2.136 | 0.0 | 30.0 | 2.072 | 1.910 | 0.0 | 30.0 |
| opposite gender | 0.137 | 0.344 | 0.0 | 1.0 | 0.044 | 0.206 | 0.0 | 1.0 |
| set 1 | 0.251 | 0.434 | 0.0 | 1.0 | 0.250 | 0.433 | 0.0 | 1.0 |
| set 2 | 0.251 | 0.433 | 0.0 | 1.0 | 0.250 | 0.433 | 0.0 | 1.0 |
| set 3 | 0.250 | 0.433 | 0.0 | 1.0 | 0.250 | 0.433 | 0.0 | 1.0 |
| set 4 | 0.249 | 0.432 | 0.0 | 1.0 | 0.249 | 0.433 | 0.0 | 1.0 |
| pairing 1 | 0.381 | 0.486 | 0.0 | 1.0 | 0.359 | 0.480 | 0.0 | 1.0 |
| pairing 2 | 0.382 | 0.486 | 0.0 | 1.0 | 0.357 | 0.479 | 0.0 | 1.0 |
| pairing 3 | 0.237 | 0.425 | 0.0 | 1.0 | 0.285 | 0.451 | 0.0 | 1.0 |
| ability | 120.414 | 9.741 | 52.0 | 149.5 | 126.289 | 9.016 | 50.2 | 154.4 |
| ability difference | -0.100 | 9.940 | -61.4 | 63.9 | 0.032 | 8.340 | -71.9 | 71.9 |
| team ability | 120.456 | 7.678 | 73.0 | 137.4 | 126.204 | 7.853 | 60.3 | 147.6 |
| opponent team ability | 120.538 | 7.652 | 61.1 | 137.4 | 126.259 | 7.812 | 63.9 | 147.6 |
| Observations | 115166 | | | | 357092 | | | |
| Distinct players | 965 | | | | 2290 | | | |

Data source: WKBV.

Table A2.2: Robustness checks: different ability measures

| | past performance | | | past & future performance | | | factors for past performance | | |
|---|---|---|---|---|---|---|---|---|---|
| | score | points | F-score | score | points | F-score | score | points | F-score |
| female | -0.079 | -0.000 | -0.051 | 0.221 | 0.000 | -0.102*** | -0.073 | -0.000 | -0.053 |
| | (0.210) | (0.006) | (0.036) | (0.153) | (0.005) | (0.029) | (0.210) | (0.006) | (0.036) |
| opp. gender | -0.097 | 0.006 | 0.002 | 0.045 | 0.003 | -0.008 | -0.099 | 0.006 | 0.002 |
| | (0.157) | (0.005) | (0.025) | (0.157) | (0.005) | (0.025) | (0.157) | (0.005) | (0.025) |
| female × opp. gender | 0.064 | -0.012 | -0.055 | -0.173 | -0.006 | -0.020 | 0.066 | -0.011 | -0.055 |
| | (0.244) | (0.008) | (0.037) | (0.231) | (0.007) | (0.035) | (0.244) | (0.008) | (0.037) |
| home | 1.564*** | 0.057*** | -0.167*** | 1.515*** | 0.053*** | -0.161*** | 1.569*** | 0.057*** | -0.168*** |
| | (0.153) | (0.004) | (0.021) | (0.143) | (0.004) | (0.019) | (0.154) | (0.004) | (0.021) |
| Constant | 22.676*** | 0.473*** | 16.223*** | -1.657 | 0.473*** | 19.667*** | 122.683*** | 0.472*** | 2.512*** |
| | (1.400) | (0.038) | (0.252) | (1.199) | (0.037) | (0.247) | (0.251) | (0.006) | (0.044) |
| Ability measure | past performance | | | past & future performance | | | factors for past performance | | |
| Observations | 64738 | 64738 | 64727 | 75056 | 75056 | 75042 | 64738 | 64738 | 64727 |
| Distinct players | 1471 | 1471 | 1471 | 1628 | 1628 | 1628 | 1471 | 1471 | 1471 |
| Adj. $R^2$ | 0.364 | 0.182 | 0.275 | 0.388 | 0.193 | 0.298 | 0.364 | 0.182 | 0.275 |

*Notes:* This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues by different types of ability measures. Models control for *pairing* and *set* (except for *points* as dependent variable). The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. *Female* and *opp. gender* are dummy variables if the player is female or plays against the opposite gender, respectively. Ability measures include own ability, difference between own and opponent's ability, the team's average ability, and the opponent team's average ability. Ability measures are: *past performance* as the average of past scores, *past & future performance* as the average of past and future scores, and *factors for past performance* from a factor analysis of past *V-score* & *A-score*. Models control for *set* and *pairing* (except for points as dependent variable), and *home*. Robust standard errors clustered at the level of the player are in parentheses. $^{***}$, $^{**}$ and $^{*}$ denote significance at the 1%, 5% and 10% level, respectively. Data source: WKBV.

Table A2.3: Robustness checks: different fixed effects

| | Individual FE | | | Lane FE | | | Individual × Lane FE | | |
|---|---|---|---|---|---|---|---|---|---|
| | score | points | F-score | score | points | F-score | score | points | F-score |
| female × opp. gender | -0.250 | -0.018** | -0.033 | 0.037 | -0.012 | -0.043 | 0.026 | -0.017 | -0.035 |
| | (0.252) | (0.008) | (0.039) | (0.235) | (0.008) | (0.037) | (0.379) | (0.013) | (0.059) |
| opp. gender | 0.002 | 0.009 | 0.010 | -0.077 | 0.006 | 0.005 | -0.004 | 0.013 | 0.021 |
| | (0.174) | (0.006) | (0.027) | (0.152) | (0.005) | (0.025) | (0.268) | (0.009) | (0.041) |
| home | 1.544*** | 0.056*** | -0.159*** | 1.555*** | 0.056*** | -0.166*** | 0.717 | 0.010 | -0.078 |
| | (0.156) | (0.004) | (0.021) | (0.130) | (0.004) | (0.020) | (0.460) | (0.015) | (0.071) |
| difference ability | -0.000 | 0.015*** | -0.002 | -0.003 | 0.015*** | -0.001 | 0.025** | 0.015*** | -0.004** |
| | (0.008) | (0.000) | (0.001) | (0.008) | (0.000) | (0.001) | (0.011) | (0.000) | (0.002) |
| female | | | | -0.061 | -0.000 | -0.061* | | | |
| | | | | (0.213) | (0.006) | (0.035) | | | |
| past ability | | | | 0.831*** | -0.000 | -0.110*** | | | |
| | | | | (0.015) | (0.000) | (0.003) | | | |
| team ability | | | | 0.028* | 0.001** | -0.006** | | | |
| | | | | (0.016) | (0.000) | (0.003) | | | |
| opponent team ability | | | | -0.035*** | -0.001*** | 0.003 | | | |
| | | | | (0.013) | (0.000) | (0.002) | | | |
| Constant | 113.114*** | 0.472*** | 3.787*** | 20.153*** | 0.473*** | 16.526*** | 113.366*** | 0.493*** | 3.746*** |
| | (0.164) | (0.003) | (0.026) | (1.530) | (0.044) | (0.271) | (0.318) | (0.008) | (0.049) |
| Fixed effects | Individual FE | | | Lane FE | | | Individual × Lane FE | | |
| Observations | 64679 | 64679 | 64679 | 63987 | 63987 | 63987 | 63987 | 63987 | 63987 |
| Distinct players | 1471 | 1471 | 1471 | 1468 | 1468 | 1468 | 1468 | 1468 | 1468 |
| Adj. $R^2$ | 0.383 | 0.195 | 0.297 | 0.384 | 0.185 | 0.284 | 0.410 | 0.200 | 0.308 |

*Notes:* This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues by different types of fixed effects. Models control for *pairing* and *set* (except for *points* as dependent variable). The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. *Female* and *opp. gender* are dummy variables if the player is female or plays against the opposite gender respectively. *Female × opp. gender Past ability* is the average *score* of the player per lane if more than 8 lanes are observable from past data. *Difference* is the difference between *past ability* of the player and her opponent. *Team ability* and *opponent team ability* are measures for team's quality calculated as the average of *past ability* of other players in the team. *Individual FE* are fixed effects for each player, *Lane FE* are fixed effects for each lane in each location, and *Individual × Lane FE* is a combination of both. Models control for *set* and *pairing* (except for *points* as dependent variable), and *home*. Robust standard errors clustered at the level of the player are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. Data source: WKBV.

Chapter 3

# Can Differences in Distributional Preferences explain the Gender Wage Gap?

# 3 Can Differences in Distributional Preferences explain the Gender Wage Gap?

*The economics department of Leibniz Universität Hannover provided a funding of 1,964 EUR. This funding body did not influence the experimental design. This funding is generally rewarded to cover implementation costs of studies in experimental economics at the Leibniz Universität Hannover.*

## 3.1 Introduction

*Пусть у меня сарай сгорит, лишь бы у соседа корова сдохла.*
*Let my shed burn down, as long as the neighbour's cow would die.*

– Russian proverb

The assumption of a self-interested *homo economicus* may be one of the most important reasons why neoclassical theory fails to predict outcomes in experiments. By dropping the assumption of narrow self-interests and accounting for other-regarding preferences one can explain many experimental outcomes, e.g. positive offers to an anonymous person in one-shot dictator games (see Engel, 2011, for an extensive meta study) and rejections of small offers in one-shot ultimatum games (Oosterbeek et al., 2004). Both of these outcomes in experimental settings are inconsistent with the neoclassical theory but can be described by distributional preferences, e.g. inequity aversion. The impact of distributional preferences outside the lab is documented by Luttmer (2005) who provides evidence that relative earnings influence self-reported well-being even after controlling for absolute earnings. The main goals of this paper are to identify subjective questions which can serve as determinants of distributional preferences and to provide indicative evidence for their impact on wages.

Subjective questions may be helpful to capture otherwise unobserved individual characteristics (Bertrand and Mullainathan, 2001). Subjective questions measure for example self-assessment of risk attitude (Dohmen et al., 2005), the Big-5 personality traits (Rammstedt and John, 2007), and self-assessment regarding the importance of career and wealth, which could serve as a proxy for distributional preferences such as inequity aversion. Previous studies confirm highly significant gender differences in other-regarding preferences (see e.g. Balafoutas et al., 2012; Güth et al., 2007). If subjective questions can serve to predict latent distributional preferences, they might reveal how these preferences affect labor market outcomes and provide an additional explanation for the gender pay gap. Following this consideration, differences in preferences of East and West Germans regarding redistribution (Alesina and Fuchs-Schündeln, 2007; Corneo, 2001; Kuhn, 2013) could provide a partial explanation for the wage gap between these two groups.

There are at least three reasons why distributional preferences might affect wages: competition, effort, and modesty. First, Balafoutas et al. (2012) and Bartling et al. (2009) show that distributional preferences affect the willingness to compete because competition by definition creates "winners" and "losers". Individuals with preferences for equality favor less competition, which in turn might be reflected by wages. Second, non-contractual effort might be less relevant for own wages due to fixed wage schemes but might be related to supervisors' earnings. If, for example, only the manager obtains a bonus for firm performance, this might create disutility for individuals with preferences for equality and might affect job performance and in the long term wages. Third, modesty considerations might be related to performance and earnings. Inequity averse persons could experience lower marginal utility from earning more than their social environment (in line with the theoretical model by Fehr and Schmidt, 1999) and hence *ceteris paribus* provide less effort to achieve this.

The contribution of this paper is to reveal the determinants of distributional preferences and analyze their influence on wages using imputed predictions of distributional preferences. The core idea is to identify which survey questions from the German Socio-Economic Panel (SOEP) can predict distributional preferences in an incentive compatible experiment. I then estimate the latent distributional preferences of SOEP-participants and show indicative evidence that inequity aversion and spiteful behavior reduce wages. Oaxaca-Blinder estimates suggest that distributional preferences account for around 3 percentage points of the wage differential between men and women.

The remainder of this paper is organized as follows: the next section describes the experiment in detail, followed by an analysis of experimental outcomes in Section 3.3. Indicative evidence on the importance of distributional preferences as wage determinants and robustness checks are presented in Section 3.4. Section 3.5 summarizes the previous results and implications of the main findings.

## 3.2 Experiment

I conducted an experiment with 118 students from Hannover. Ten sessions ran from October to November 2013. 40% of the participants were Economics and Business students, 43% enrolled in STEM fields[1]. A heterogeneous sample is desirable since students who self-select in economics and business studies may differ in their distributional preferences and taste for efficiency compared to students of other fields. The experiment consists of two parts.[2] The first part is a paper-based questionnaire on the socioeconomic status and a variety of self-assessment items, some taken from the SOEP (Goebel et al., 2019). The second part consists of a computer-based version of the incentivized *Equality Equivalence Test* on distributional preferences by Kerschbamer (2015) implemented in *z-tree* (Fischbacher, 2007).

The *Equality Equivalence Test* reveals how much an individual in advantageous and disadvantageous situations is willing to pay for an increase in payment to an unknown person. This test is simple and new, which makes it unlikely that it is known to students which would bias the estimates. The *Equality Equivalence Test* is a one-shot experiment excluding any strategic considerations. It focuses on distributional preferences only and abstracts from reciprocity, peer effects, trust attitudes, and preferences for cooperation.

Participants choose between two alternatives in each of the ten questions (see table 3.1): *Option LEFT* or *Option RIGHT*. Each alternative provides a pair of payoffs, one for the decision maker and one for an anonymous person in the room, who is called *passive person* because she has no part in the respective decision. All participants answer the full set of questions in a random order. Participants see only one question at a time and cannot return to previous decisions. This allows identifying and excluding persons who click randomly, who do not understand the experiment or have no clear opinion about their distributional preferences. Once the session is completed, 50% of the participants are chosen randomly to be a *passive person*, so that their previous decisions have no influence on their payoff.[3] Based on one randomly chosen question, active participants are paid according to their decision, while passive participants receive the payment assigned by their active counterpart.

[1] Science, Technology, Engineering, and Mathematics.
[2] There was a third part on measuring risk preferences which is not used in this paper.
[3] Hedegaard et al. (2011) provide evidence that the result of the *Equality Equivalence Test* with random role assignment does not significantly differ from a fixed role implementation.

Table 3.1: Incentivised questions of the *Equality Equivalence Test*

| | Option LEFT | | Option RIGHT | |
|---|---|---|---|---|
| | You get: | Passive person gets: | You get: | Passive person gets: |
| 1 | 3.20 EURO | 2.80 EURO | 4.00 EURO | 4.00 EURO |
| 2 | 3.60 EURO | 2.80 EURO | 4.00 EURO | 4.00 EURO |
| 3 | 4.00 EURO | 2.80 EURO | 4.00 EURO | 4.00 EURO |
| 4 | 4.40 EURO | 2.80 EURO | 4.00 EURO | 4.00 EURO |
| 5 | 4.80 EURO | 2.80 EURO | 4.00 EURO | 4.00 EURO |
| 6 | 3.20 EURO | 5.20 EURO | 4.00 EURO | 4.00 EURO |
| 7 | 3.60 EURO | 5.20 EURO | 4.00 EURO | 4.00 EURO |
| 8 | 4.00 EURO | 5.20 EURO | 4.00 EURO | 4.00 EURO |
| 9 | 4.40 EURO | 5.20 EURO | 4.00 EURO | 4.00 EURO |
| 10 | 4.80 EURO | 5.20 EURO | 4.00 EURO | 4.00 EURO |

*Option RIGHT* provides the same payoff for decision maker and *passive person* and remains unchanged in all 10 questions. The first five questions in table 3.1 are called the advantageous decision block as all payoffs from *Option LEFT* are advantageous to the decision maker. The last five questions are the disadvantageous decision block. A rational individual should not switch more than once from *Option RIGHT* to *Option LEFT* in each of the two blocks and never switch in the opposite direction (assuming that the marginal utility of own payoff is always positive). 20 participants (17%) with a switching point in an irrational direction are excluded from the further analysis.[4] Since the occurrence of such an irrational decision pattern is not correlated with gender, age, the instructional manipulation check (Oppenheimer et al., 2009), and the cognitive reflection test (Frederick, 2005), there is no evidence for systematic exclusions.

Willingness to pay ($WTP$) is the amount an individual is willing to pay for a €1 gain or loss of an anonymous person. If, for example, a participant is willing to forego €0.30 when the *passive person* gains €1, the $WTP$ would be 0.3. The switching point in the first block reveals the $WTP$ in an advantageous situation ($WTPADV$) of the decision maker. For example, if one person chooses *Option RIGHT* in the first four questions and *Option LEFT* in the fifth question,

[4]This proportion is lower than the 27% in the online experiment by Hedegaard et al. (2011), but higher than in experiments which offer an ordered summary of the decisions made and allow to revise previous decisions (8% in Balafoutas et al. (2012); 4% in Kerschbamer (2015)).

she is willing to give up €0.40 for a payment increase of €1.20 for a *passive person* but is not willing to give up €0.80 for the same increase. As a consequence, *WTPADV* can be narrowed down to between $1/3$ and $2/3$ and is coded as 0.5, the mean of these two boundaries.

The willingness to pay in a disadvantageous situation (*WTPDIS*) is calculated similarly. For example if the individual chooses *Option RIGHT* in question 6 it would reveal that the person is not willing to forego €0.80 for a payment increase of €1.20 for a *passive person*, so that *WTPDIS* is smaller than $2/3$. Hence if the decision maker choses *Option RIGHT* in the questions 6 , 7, 8, and 9, while switching in question 10 to *Option LEFT* *WTPDIS* can be narrowed down to between $-1/3$ and $-2/3$ and is coded as $-0.5$. The analysis is done with the full range of *WTP*.[5]

Kerschbamer (2015) classifies distributional preferences into four types:

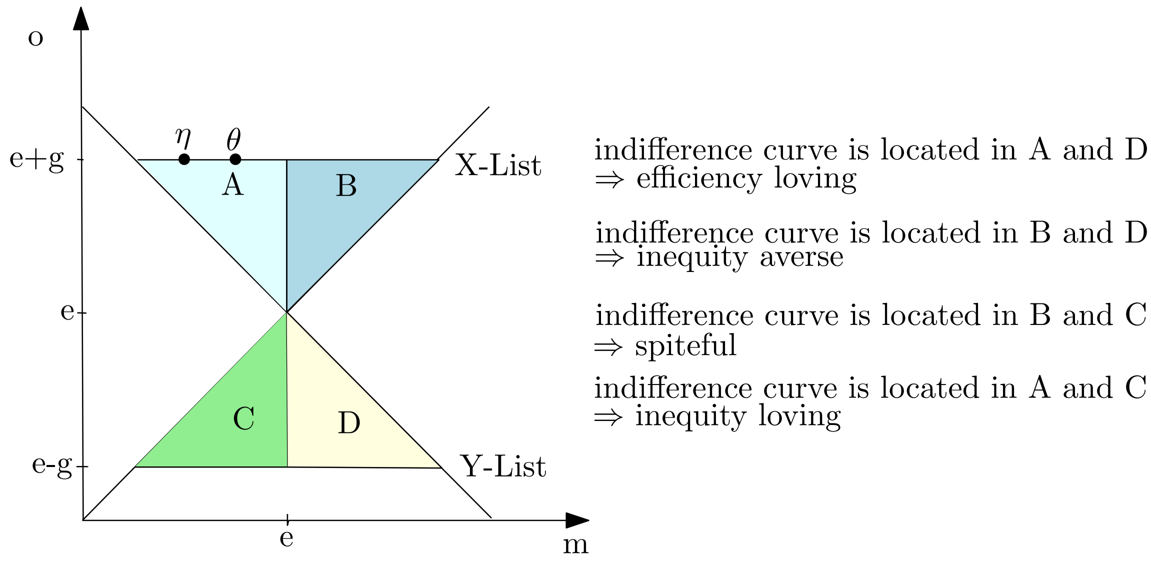
| | |
|---|---|
| "Efficiency loving " (ELO): | WTPADV$>$ 0, WTPDIS$>$ 0 |
| "Inequity averse" (IAV): | WTPADV$>$ 0, WTPDIS$<$ 0 |
| "Spiteful" (SPI): | WTPADV$<$ 0, WTPDIS$<$ 0 |
| "Inequity loving" (ILO) : | WTPADV$<$ 0, WTPDIS$>$ 0 |

This paper focuses on the difference between the efficient (*ELO* preferences) and inefficient (*SPI* and *IAV*) level of effort. This difference is captured by the sign of *WTPDIS*. Subjects with *ELO* preferences have a positive *WTP* in both environments and are up to a certain degree willing to increase the sum of payments for themselves and a *passive person*. Negative *WTPDIS* results in an inefficient effort level in the labor market because individuals are willing to pay money in order to decrease other's payoff, e.g. of their supervisor. $WTPDIS < 0$ holds for *SPI* and *IAV* preferences. *ILO* preferences are only defined for completeness.

Figure 3.1 provides the graphical representation of the *Equality Equivalence Test*. Along the indifference curve the decision maker is indifferent between own payoff $m$ and the payoff of an anonymous *passive person* $o$. *X-List* are the five questions in the disadvantageous situation (with $m < o$) and *Y-List* the advantageous situation (with $m > o$). The equal payoff corresponds to the *Option RIGHT* and is denoted by $e$ (€4 in this experiment). $g$ is the difference between the *passive person*'s payoff in *Option RIGHT* and *Option LEFT* (€1.20 in this experiment). If the individual switches between question $\eta$ and $\theta$ from *Option RIGHT* to *Option LEFT*, this narrows down the location of her indifference curve between these two points.

[5]Probit results for the binary outcomes $WTP > 0$ and results for Tobit estimation are qualitatively similar and available upon request.

Figure 3.1: *Equality Equivalence Test* (indifference curves representation)




*Notes:* Own adaptation of (Kerschbamer, 2015, fig. 4, page 95). *m* denotes the the payoff of the decision maker and *o* is the payment for the passive person. Equal payment for decision maker and passive person is denoted by *e* and represents *Option RIGHT* from table 3.1. *X-List* are the five questions in the disadvantageous situation (with $m < o$) and *Y-List* the advantageous situation (with $m > o$). $\eta$ and $\theta$ denote exemplary payoff combinations for *Option LEFT* from the disadvantageous situation.

.

## 3.3 Determinants of distributional preferences

This section summarizes gender differences in distributional preferences and presents determinants of *WTPDIS* from the experiment. The estimated coefficients are used in section 3.4 in order to predict distributional preferences for SOEP participants. Consistent with previous experiments (Andreoni and Vesterlund, 2001; Eckel and Grossman, 1998), I verify that men are more often efficiency loving (*ELO*) and women are more often inequity averse (*IAV*), see figure 3.2 and figure 3.3. Men are more often willing to maximize the sum of payments between themselves and the passive person. They are less inclined to pay money for value destruction, i.e. the passive person receiving less. Both results are significant at the 1% level with Pearson's $\chi^2$ and Fisher's exact test. The largest variation between the genders occurs in *WTPDIS* (Fig. 3.3), so there are no significant differences in *WTPADV* between genders (see Appendix figure A3.1). Besides participants gender also other variables cannot explain variation in *WTPADV* (see Appendix table A3.1).

Figure 3.2: *Equality Equivalence Test* (Distributional preferences by gender)

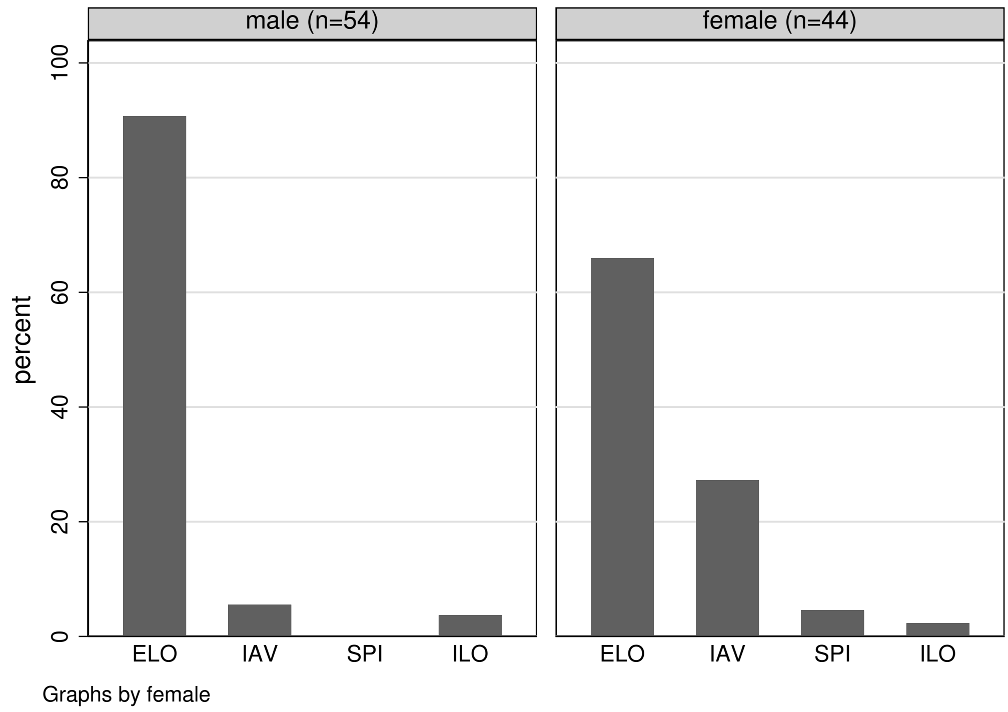
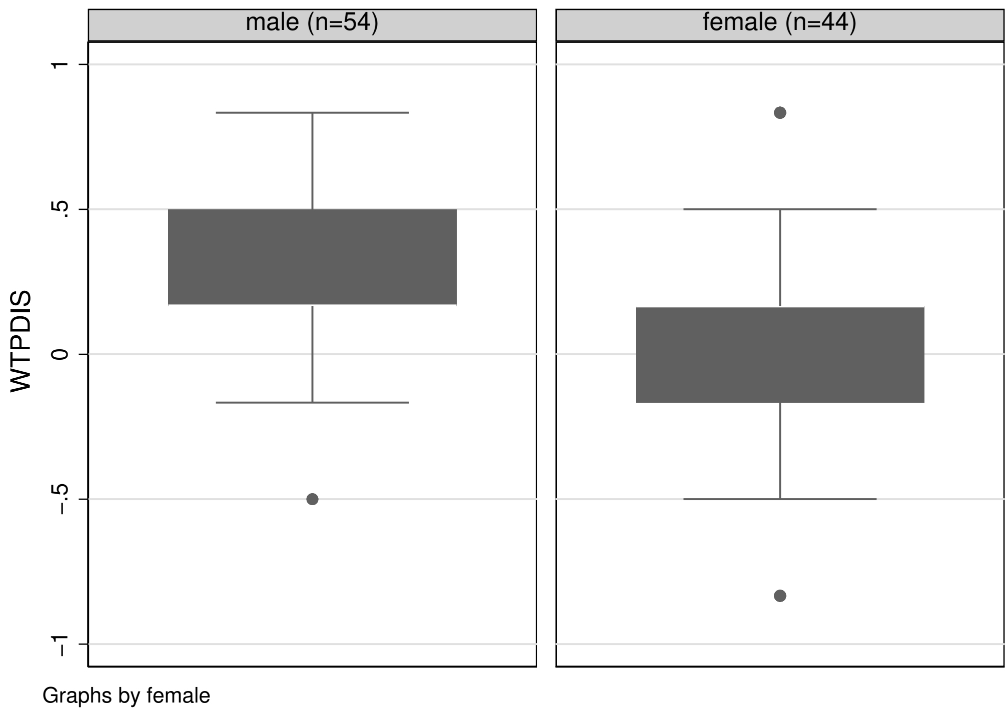



Figure 3.3: WTPDIS by gender



I estimate the willingness to pay in a disadvantageous environment $WTPDIS_i$ according to equation (3.1), where $impsuccess_i$ is the importance of job success for individual $i$ (a SOEP item measured on a 4-point scale, ranging from very important to very unimportant) and $peduc_i$ is a dummy variable indicating if one of the parents had at least qualification for university entrance (*Ger. Abitur*). $female_i$ is a dummy variable indicating gender, $X_i'$ is a vector containing Big-5 personality measures and risk preferences, and $\varepsilon_i$ are the heteroscedasticity robust error terms.[6]

$$WTPDIS_i = \beta_0 + \beta_1 impsuccess_i + \beta_2 peduc_i + \beta_3 female_i + X_i'\gamma + \varepsilon_i \tag{3.1}$$

Table 3.2 shows the determinants of distributional preferences. A good predictor for $WTPDIS$ is the self-reported importance of success at work. Subjects who claim that career is not important for them are more likely to have positive $WTPDIS$, i.e. less like to be inequity averse. Parental education is associated with higher $WTPDIS$. Agreeableness and conscientiousness of the Big-5 personality traits are negatively related to $WTPDIS$.[7] Gender is an important determinant for distributional preferences even controlling for personality traits. The coefficient is significant at the 1% level and suggests that women are willing to pay 25 cents less for a € 1-increase of the *passive person*'s payment. Self-assessed risk measures are significant in some specifications.[8] Robustness checks include controls for CRT (Cognitive Reflection Test, see Frederick, 2005)[9], IMC (Instructional Manipulation Check, see Oppenheimer et al., 2009)[10], number of siblings, and political preferences and are reported in the Appendix, table A3.2.

[6]The results are qualitatively similar with Tobit (since $WTP$ is limited between -1 and 1) and Probit models (limited to a positive $WTPDIS$). Estimating $WTPADV$ does not reveal significant determinants (see Appendix table A3.1).

[7]This is partially in line with the results of Hedegaard et al. (2011), who found that agreeableness and openness are significant determinants for contribution in a public good experiment.

[8]Subjects were asked about their self-reported willingness to take risk in different spheres: general, driving, financial matters, leisure, career, and health. General risk attitude and risk attitude towards financial matters are significantly related to $WTPDIS$.

[9]The test consists of three questions each of those has an intuitive but incorrect answer. The right answer requires conscious thought and is therefore attributed to system 2 decision making.

[10]This is a special question in order to check whether participants read the instructions or respond randomly. Using the same layout as in the remaining question the participants are asked to select two specific items on a Likert scale.

Table 3.2: Determinants of distributional preferences

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| importance: success work | -0.194*** | -0.215*** | -0.235*** | -0.238*** | -0.224*** |
| | (0.063) | (0.064) | (0.061) | (0.064) | (0.062) |
| parents A-Level | | 0.154*** | 0.140*** | 0.116** | 0.133** |
| | | (0.053) | (0.050) | (0.057) | (0.055) |
| openness | | | 0.010 | 0.002 | 0.021 |
| | | | (0.037) | (0.037) | (0.038) |
| extraversion | | | 0.053 | 0.033 | 0.025 |
| | | | (0.038) | (0.045) | (0.043) |
| neuroticism | | | -0.077** | -0.055 | -0.030 |
| | | | (0.035) | (0.038) | (0.038) |
| conscientiousness | | | -0.073** | -0.085** | -0.065* |
| | | | (0.036) | (0.038) | (0.039) |
| agreeableness | | | -0.085* | -0.103** | -0.104** |
| | | | (0.043) | (0.048) | (0.048) |
| female | | | | | -0.254*** |
| | | | | | (0.087) |
| Constant | 0.777*** | 0.666*** | 0.729*** | 0.571** | 0.744*** |
| | (0.191) | (0.196) | (0.178) | (0.274) | (0.267) |
| Risk [Self Assessment] | – | – | – | √ | √ |
| Observations | 98 | 98 | 98 | 97 | 97 |
| Adj. $R^2$ | 0.093 | 0.163 | 0.218 | 0.237 | 0.286 |

*Notes:* This table shows the relationship between distributional preferences and self-assessment question on sociodemographic characteristics and personality traits. The outcome is willingness to pay in a disadvantageous situation, denoted by *WTPDIS*. *importance: success work* is a self assessed question on importance of success of work, ranging from 1 (very unimportant) to 4 (very important). *parents A-level* and *female* are a dummy variables if one of the parents had an A-level certificate and if the individual is female. *extraversion*, *neuroticism*, *openness*, *conscientiousness*, and *agreeableness* denote the standardized Big-5 personality traits. Robust standard errors in parentheses. ***, **, and * denote significance with at the 1%, 5%, and 10% level, respectively.

Even though there are uncountable possibilities between minus one and one, neoclassical theory assumes a $WTP$ of exactly zero. If $WTP = 0$ occurred systematically, the share of people with negative $WTP$ would be overestimated and these individuals would choose lower payment for a passive person with a 50% chance. In this situation the dependent variable from table 3.2 would provide an upper bound for individuals with strictly positive $WTPDIS$. The implications remain unchanged as a higher $WTPDIS$ is associated with more efficient distributional preferences and should be rewarded in the labor market.

## 3.4 Analyzing the Gender wage gap

To address whether distributional preferences are related to earnings variation in Germany, I predict $WTPDIS$ for participants of a representative survey. I construct $WTPDIS$ of SOEP-participants using the coefficients from the experiment as described in the previous section and show that predicted distributional preferences are related to lower wages and can partly explain the gender pay gap in the Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca, 1973; Jann, 2008).

I cannot make any causal claims because of two reasons: external validity and direct effects. The first concern is the external validity of the experimental results, i.e. that the student sample from the experiment should be representative of SOEP-participants in the same age group (18 to 32). Narrowing the analysis to those holding a university degree and who answered the relevant personality questions between 18 and 32 year of age confirms that the magnitude of the effect is slightly larger compared to the whole sample of young adults.

The second concern is the direct effect of self-assessed items such as Big-5 and importance of success at work on earnings. If there are such direct effects, a significant coefficient for $WTPDIS$ would imply a correlation but not a causal effect (similar to violations of the exclusion restriction in the Two Stage Least Squares approach). One reason could be reverse causality, e.g. between importance of success at work and earnings. If this is the case, one would expect that individuals with higher wages state that success is important to them. In the experiment, I show that high importance of success is associated with lower $WTPDIS$. This is in turn related to lower wages suggesting that importance of success at work is associated with lower wages. Hence, if there are direct effects, these would bias the coefficients of $\widehat{WTPDIS}$ towards zero. Concerning the potential direct effect of Big-5 personality traits on earnings (see e.g. Mueller and Plug, 2006, who analyse gender-specific returns to personality traits), I show that positive $WTPDIS$ is still related to higher wages even if controlling for Big-5 personality traits in some specifications.

For the analysis I transform the hourly wage with the natural logarithm following the skewness of the distribution. I limit the analysis to SOEP-participants employed full-time, with earnings of more than €3000 per year who answered all relevant personality questions in the same range of age as the students from the experiment (18 to 32).[11] Implicitly, I assume that $WTPDIS$ is stable for young adults. This is supported by Fehr et al. (2013) showing that distributional preferences develop during adolescence and are stable in adulthood, with a significant gap in egalitarian preferences between the genders. I calculate $WTPDIS$ according to equation (3.1) using the estimated coefficients of column (3) from table 3.2. The predicted values ($\widehat{WTPDIS}$) are used as an explanatory variable. I regress the natural logarithm of the hourly wage of individual $i$ on a constant $\alpha$, the predicted distributional preferences ($\widehat{WTPDIS}$), and a vector of sociodemographics and work characteristics $Z_i'$. $u_i$ denotes the error term clustered on individual level. Bootstrapped standard errors are reported as a robustness check in squared brackets. I estimate equation (3.2) with pooled OLS.

$$\ln(hwage_i) = \alpha + \beta \widehat{WTPDIS}_i + Z_i'\eta + u_i \tag{3.2}$$

There is a significant positive relationship between $\widehat{WTPDIS}$ and hourly wages in all specifications in tables 2.3a and 2.3b. The coefficient ranges from 0.341 in a model including year fixed effects only to 0.073 in the full specification. Being willing to pay 10 cents more for each €1 in the *passive person*'s payoff in a disadvantageous situation is associated with 0.73 to 3.41% higher wages. Augmenting the parsimonious specification from column (1) in table 2.3a by sociodemographic and work characteristics reduces the coefficient of interest. Occupational choice and regional differences might explain this reduction as e.g. Alesina and Fuchs-Schündeln (2007) document that East and West Germans differ in their preferences for redistribution. In column (6) of table 2.3a control variables for Big-5 personality traits are added, which decreases the coefficient for $\widehat{WTPDIS}$ from 0.14 to 0.109, but remains significant at the 1% level. Adding controls for educational degree in column (8) of table 2.3b substantially decreases the coefficient of $\widehat{WTPDIS}$ from 0.142 to 0.069, probably due to self-selection into school track which is associated with differences in distributional preferences (see e.g John and Thomsen, 2015, for evidence

[11] Individuals older then 32 can appear in the sample if they answered the relevant questions until the age of 32.

from Germany). Controlling for types of leadership position in column (10) considerably reduces sample size to one third as this item is not routinely included in the SOEP questionnaire. The coefficient of interest remains quite stable but is significant at the 5% level only.

The correct Two-Sample Two-Stage Least Squares (TS2SLS) standard errors would take into account that the variable $WTPDIS$ is estimated but not observed for SOEP-participants. Pacini and Windmeijer (2016) propose an approach to adjust the standard errors for the second stage, which requires that all control variables being observable in both samples. Table A3.3 compares the results of the second stage using the reduced set of covariates in column (1), the adapted Stata code from Pacini and Windmeijer (2016) is in the appendix 3.1 on page 47. Clustered standard errors in column (2) are qualitatively similar, while bootstrapped standard errors in columns (3) and (4) are lower. Estimates in column (5) report the coefficient for $WTPDIS$ estimated without Big-5 in the first stage and controlling for them in the second stage. The coefficient of $WTPDIS$ is smaller in magnitude but is still significant at the 5%-level. Hence, using clustered standard errors leads, if at all, to larger errors and larger p-values.

In order to analyze whether distributional preferences can partly explain the gender pay gap I use the Oaxaca-Blinder decomposition. Differences in $\widehat{WTPDIS}$ can explain 1.5 to 5 percentage points in the gender wage gap. Table 3.4 displays the Oaxaca-Blinder decomposition for selected specifications as indicated by the column numbers. Adding control variables reduces the contribution of distributional preferences to explain the gender wage gap. In the full specification in column (12), 6 percentage points remain unexplained. $\widehat{WTPDIS}$ still explains 2.2 percentage points but this is significant at the 5% level only.

Table 2.3a: Determinants of wages

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| $\widehat{WTPDIS}$ | 0.341*** | 0.242*** | 0.215*** | 0.180*** | 0.140*** | 0.109*** |
| | (0.027) | (0.030) | (0.030) | (0.028) | (0.026) | (0.038) |
| | [0.014] | [0.014] | [0.015] | [0.015] | [0.013] | [0.019] |
| female | | -0.012 | -0.031* | -0.047*** | -0.045*** | -0.034** |
| | | (0.018) | (0.018) | (0.017) | (0.015) | (0.017) |
| | | [0.008] | [0.009] | [0.007] | [0.010] | [0.010] |
| nchild | | 0.012* | 0.016** | 0.014** | 0.015*** | 0.017*** |
| | | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| | | [0.003] | [0.003] | [0.003] | [0.003] | [0.003] |
| female × nchild | | -0.094*** | -0.120*** | -0.065*** | -0.064*** | -0.065*** |
| | | (0.013) | (0.014) | (0.014) | (0.012) | (0.012) |
| | | [0.007] | [0.007] | [0.007] | [0.009] | [0.007] |
| age | | 0.065*** | 0.075*** | 0.080*** | 0.069*** | 0.069*** |
| | | (0.011) | (0.011) | (0.010) | (0.009) | (0.009) |
| | | [0.007] | [0.009] | [0.007] | [0.007] | [0.009] |
| age$^2$ | | -0.001*** | -0.001*** | -0.001*** | -0.000*** | -0.000*** |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| | | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| working time | | | -0.012*** | -0.011*** | -0.010*** | -0.010*** |
| | | | (0.001) | (0.001) | (0.001) | (0.001) |
| | | | [0.001] | [0.001] | [0.000] | [0.000] |
| $\mathrm{exp}_{\mathrm{ft}}$ | | | -0.014*** | -0.014*** | -0.012*** | -0.012*** |
| | | | (0.002) | (0.002) | (0.002) | (0.002) |
| | | | [0.001] | [0.001] | [0.001] | [0.001] |
| $\mathrm{exp}_{\mathrm{pt}}$ | | | -0.021*** | -0.026*** | -0.022*** | -0.021*** |
| | | | (0.004) | (0.004) | (0.004) | (0.004) |
| | | | [0.002] | [0.002] | [0.002] | [0.002] |
| State-FE | – | – | – | √ | √ | √ |
| Firm size | – | – | – | – | √ | √ |
| Big-5 | – | – | – | – | – | √ |
| Observations | 24,652 | 24,652 | 24,652 | 24,652 | 24,652 | 24,652 |
| Adj. $R^2$ | 0.059 | 0.140 | 0.174 | 0.258 | 0.327 | 0.332 |

*Notes:* This table shows the relationship between the natural logarithm of hourly labor earnings and distributional preferences. $\widehat{WTPDIS}$ denotes the estimated willingness to pay in a disadvantageous situation. *nchild* are the number of children under 16 in the household. *working time* are the average weekly working hours. $exp_{ft}$ and $exp_{pt}$ denote the labor market experience in full-time and part-time occupations, respectively. All regressions include year fixed effects. *Firm size*, *State-FE* denote fixed effects for firm size and for state of residence, respectively. *Big-5* indicate control variables for Big-5 personality traits. Robust clustered (by individual) standard errors in parentheses and bootstrapped standard errors are reported in squared brackets. ***, **, and * denote significance with the highest standard errors at the 1%, 5%, and 10% level, respectively.

Table 2.3b: Determinants of wages, continued

| | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|
| $\widehat{WTPDIS}$ | 0.142*** | 0.069*** | 0.054** | 0.070** | 0.072** | 0.073** |
| | (0.026) | (0.023) | (0.022) | (0.033) | (0.033) | (0.033) |
| | [0.013] | [0.011] | [0.013] | [0.020] | [0.023] | [0.024] |
| female | -0.043*** | -0.109*** | -0.104*** | -0.075*** | -0.073*** | -0.075*** |
| | (0.015) | (0.014) | (0.014) | (0.020) | (0.020) | (0.020) |
| | [0.007] | [0.008] | [0.009] | [0.013] | [0.015] | [0.010] |
| nchild | 0.006 | 0.010* | 0.009* | 0.003 | 0.004 | 0.008 |
| | (0.006) | (0.005) | (0.005) | (0.008) | (0.008) | (0.008) |
| | [0.003] | [0.003] | [0.003] | [0.005] | [0.005] | [0.006] |
| female $\times$ nchild | -0.058*** | -0.014 | -0.007 | 0.006 | 0.006 | 0.007 |
| | (0.012) | (0.011) | (0.011) | (0.016) | (0.016) | (0.016) |
| | [0.006] | [0.006] | [0.006] | [0.009] | [0.011] | [0.010] |
| age | 0.066*** | 0.016* | 0.014 | 0.016 | 0.011 | 0.009 |
| | (0.010) | (0.009) | (0.009) | (0.018) | (0.018) | (0.018) |
| | [0.008] | [0.008] | [0.007] | [0.015] | [0.016] | [0.016] |
| age$^2$ | -0.000*** | -0.000 | -0.000 | -0.000 | -0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| working time | -0.010*** | -0.012*** | -0.011*** | -0.014*** | -0.014*** | -0.014*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| | [0.001] | [0.000] | [0.000] | [0.001] | [0.001] | [0.001] |
| $exp_{ft}$ | -0.012*** | 0.013*** | 0.013*** | 0.012*** | 0.010*** | 0.010*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| | [0.001] | [0.001] | [0.001] | [0.001] | [0.002] | [0.002] |
| $exp_{pt}$ | -0.021*** | -0.013*** | -0.011*** | -0.010*** | -0.009** | -0.009** |
| | (0.004) | (0.003) | (0.003) | (0.004) | (0.004) | (0.004) |
| | [0.002] | [0.002] | [0.002] | [0.002] | [0.002] | [0.003] |
| Education | – | √ | √ | √ | √ | √ |
| Industry | – | – | √ | √ | √ | √ |
| Leadership position | – | – | – | √ | √ | √ |
| Career breaks | – | – | – | – | √ | √ |
| Immigration status | – | – | – | – | – | √ |
| Observations | 24,652 | 24,652 | 24,652 | 8,744 | 8,744 | 8,744 |
| Adj. $R^2$ | 0.331 | 0.416 | 0.444 | 0.468 | 0.473 | 0.477 |

*Notes:* This table shows the relationship between the natural logarithm of hourly labor earnings and distributional preferences. $\widehat{WTPDIS}$ denotes the estimated willingness to pay in a disadvantageous situation. *nchild* are the number of children under 16 in the household. *working time* are the average weekly working hours. $exp_{ft}$ and $exp_{pt}$ denote the labor market experience in full-time and part-time occupations, respectively. All regressions include fixed effects for year, state, firm size, and marital status. *Education*, *Industry*, *Leadership position*, *Career breaks*, and *Immigration status* denote fixed effects for highest education degree, 2-digit industry code, managerial responsibility, amount of job changes and career breaks, and duration of residence in Germany. Robust clustered (by individual) standard errors in parentheses and bootstrapped standard errors are reported in squared brackets. ***, **, and * denote significance with the highest standard errors at the 1%, 5%, and 10% level, respectively.

Table 3.4: Oaxaca Blinder results

| | (1) | (2) | (3) | (4) | (5) | (7) | (9) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|
| overall | | | | | | | | | |
| male ln(hwage) | 2.692*** | 2.692*** | 2.692*** | 2.692*** | 2.692*** | 2.692*** | 2.692*** | 2.738*** | 2.738*** |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.008) | (0.008) | (0.008) | (0.012) | (0.012) |
| female ln(hwage) | 2.507*** | 2.507*** | 2.507*** | 2.507*** | 2.507*** | 2.507*** | 2.507*** | 2.588*** | 2.588*** |
| | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.014) | (0.014) |
| difference | 0.185*** | 0.185*** | 0.185*** | 0.185*** | 0.185*** | 0.185*** | 0.185*** | 0.150*** | 0.150*** |
| | (0.013) | (0.013) | (0.013) | (0.013) | (0.013) | (0.013) | (0.013) | (0.018) | (0.018) |
| explained | 0.047*** | 0.108*** | 0.054*** | 0.077*** | 0.079*** | 0.084*** | 0.080*** | 0.091*** | 0.089*** |
| | (0.010) | (0.012) | (0.013) | (0.014) | (0.014) | (0.013) | (0.014) | (0.021) | (0.020) |
| unexplained | 0.138*** | 0.077*** | 0.131*** | 0.108*** | 0.106*** | 0.101*** | 0.105*** | 0.059*** | 0.062*** |
| | (0.016) | (0.017) | (0.018) | (0.017) | (0.016) | (0.016) | (0.016) | (0.022) | (0.022) |
| explained | | | | | | | | | |
| $\widehat{WTPDIS}$ | 0.040*** | 0.050*** | 0.037*** | 0.033*** | 0.022*** | 0.022*** | 0.015** | 0.023** | 0.022** |
| | (0.010) | (0.010) | (0.009) | (0.009) | (0.008) | (0.008) | (0.007) | (0.011) | (0.011) |
| unexplained | | | | | | | | | |
| $\widehat{WTPDIS}$ | 0.018** | 0.018** | 0.026*** | 0.019** | 0.020*** | 0.021*** | 0.018*** | 0.013 | 0.016* |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.007) | (0.007) | (0.007) | (0.009) | (0.009) |
| Observations | 24,652 | 24,652 | 24,652 | 24,652 | 24,652 | 24,652 | 24,652 | 8,744 | 8,744 |

*Notes:* This table shows the Oaxaca Blinder estimates (Jann, 2008) for selected regression model described in Tables 2.3a & 2.3b. $\widehat{WTPDIS}$ denotes the estimated willingness to pay in a disadvantageous situation. Robust clustered (by individual) standard errors are in parentheses. ***, **, and * denote significance with the highest standard errors at the 1%, 5%, and 10% level, respectively.

## 3.5 Conclusion

Neoclassical theory neglects other-regarding preferences whose existence is documented in several laboratory experiments. I make a first step into showing how distributional preferences relate to earnings. Measuring distributional preferences in an experiment and developing a model to predict them based on personality traits, I construct the distributional preferences of SOEP-participants and provide indicative evidence for their role for earnings and the gender pay gap.

Being willing to pay 10 cents more for each € 1 to an unknown person is associated with an increase in wages of 0.54 to 3.41%. Differences in distributional preferences can explain 1.5 to 5 percentage points in the gender wage gap. In line with theoretical considerations and previous literature I find that the part explained by distributional preferences declines due to sorting into education, occupations, and leadership positions. These sorting patterns can be also regarded as endogenous, as they might be the consequence of distributional preferences and therefore lead to a downward bias of the coefficient of interest. Nonetheless coefficient of other-regarding preferences on wages remains significant. This suggests that ignoring differences in distributional preferences between men and women can lead to a biased estimate of the pay gap.

Using imputed distributional preferences provides only indicative evidence as relying on strong assumptions. First, to achieve external validity, the student sample from the experiment needs to be representative of the SOEP-participants in the same age group. Second, direct effects of the predictors of distributional preferences might bias the estimates, e.g. direct effects of the Big-5 personality traits. Hence, estimated distributional preferences of SOEP-participants provide only an imperfect proxy for their other-regarding preferences. Future research should address these drawbacks and preferably relate earning directly to measured distributional preferences.

## 3.6 Appendix

Figure A3.1: WTPADV by gender

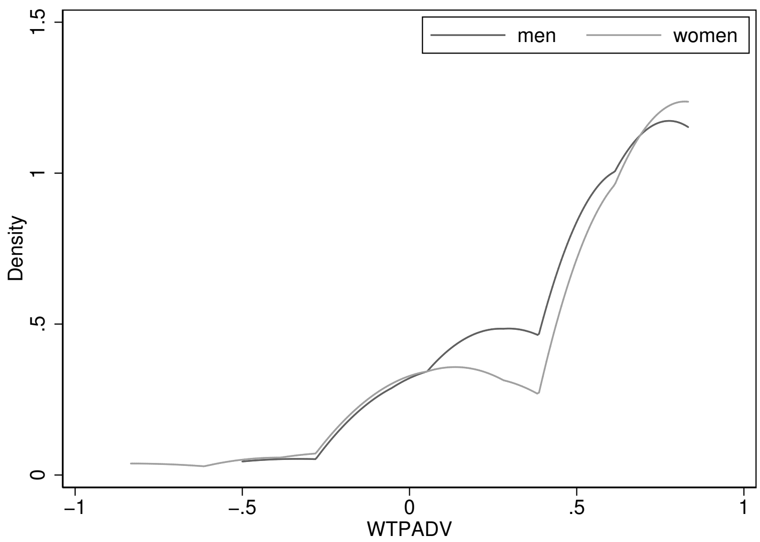


Table A3.1: Determinants of distributional preferences (WTPADV)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| importance: success work | -0.062 | -0.067 | -0.077 | -0.063 | -0.062 |
| | (0.046) | (0.045) | (0.056) | (0.055) | (0.054) |
| parents A-Level | | 0.034 | 0.025 | 0.027 | 0.029 |
| | | (0.052) | (0.052) | (0.055) | (0.054) |
| openness | | | -0.020 | -0.021 | -0.020 |
| | | | (0.041) | (0.038) | (0.037) |
| extraversion | | | 0.025 | 0.032 | 0.032 |
| | | | (0.044) | (0.049) | (0.048) |
| neuroticism | | | -0.034 | -0.072 | -0.070 |
| | | | (0.035) | (0.047) | (0.048) |
| conscientiousness | | | 0.000 | -0.020 | -0.019 |
| | | | (0.037) | (0.054) | (0.054) |
| agreeableness | | | 0.009 | 0.003 | 0.003 |
| | | | (0.039) | (0.045) | (0.046) |
| female | | | | | -0.021 |
| | | | | | (0.114) |
| Constant | 0.806*** | 0.782*** | 0.819*** | 1.004*** | 1.019*** |
| | (0.133) | (0.149) | (0.183) | (0.273) | (0.309) |
| Risk [Self Assessment] | – | – | – | √ | √ |
| Observations | 98 | 98 | 98 | 97 | 97 |
| Adj. $R^2$ | 0.005 | -0.001 | -0.036 | -0.056 | -0.069 |

*Notes:* This table shows the relationship between distributional preferences and self-assessment question on sociodemographic characteristics and personality traits. The outcome is willingness to pay in a advantageous situation (*WTPADV*). *importance: success work* is a self assessed question on importance of success of work, ranging from 1 (very unimportant) to 4 (very important). *parents A-level* and *female* are a dummy variables if one of the parents had an A-level certificate and if the individual is female. *extraversion*, *neuroticism*, *openness*, *conscientiousness*, and *agreeableness* denote the standardized Big-5 personality traits. Robust standard errors in parentheses. ***, **, and * denote significance with at the 1%, 5%, and 10% level, respectively.

Table A3.2: Determinants of distributional preferences (robustness checks)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| importance: success work | -0.224*** | -0.211*** | -0.212*** | -0.213*** | -0.209*** | -0.207*** |
| | (0.062) | (0.063) | (0.061) | (0.062) | (0.060) | (0.061) |
| parents A-Level | 0.133** | 0.133** | 0.130** | 0.127** | 0.106* | 0.123** |
| | (0.055) | (0.055) | (0.055) | (0.055) | (0.057) | (0.060) |
| female | -0.254*** | -0.250*** | -0.241** | -0.248** | -0.258*** | -0.266** |
| | (0.087) | (0.092) | (0.091) | (0.095) | (0.097) | (0.101) |
| IMC (fail) | | 0.133 | 0.102 | 0.108 | 0.114 | 0.155 |
| | | (0.094) | (0.097) | (0.095) | (0.095) | (0.102) |
| crtsuccess | | | 0.058 | 0.055 | 0.047 | 0.037 |
| | | | (0.040) | (0.040) | (0.041) | (0.045) |
| siblings | | | | 0.020 | 0.026 | 0.025 |
| | | | | (0.047) | (0.046) | (0.045) |
| grade A-level | | | | | -0.081 | -0.072 |
| | | | | | (0.067) | (0.067) |
| politics left-right | | | | | | 0.012 |
| | | | | | | (0.019) |
| Constant | 0.744*** | 0.576* | 0.480 | 0.466 | 0.665** | 0.550* |
| | (0.267) | (0.305) | (0.304) | (0.301) | (0.309) | (0.302) |
| Big-5 | √ | √ | √ | √ | √ | √ |
| Risk [Self Assessment] | √ | √ | √ | √ | √ | √ |
| Observations | 97 | 96 | 96 | 96 | 96 | 95 |
| Adj. $R^2$ | 0.286 | 0.276 | 0.290 | 0.283 | 0.285 | 0.291 |

*Notes:* This table shows the relationship between distributional preferences and self-assessment question on sociodemographic characteristics and personality traits. The outcome is willingness to pay in a disadvantageous situation, denoted by *WTPDIS*. *importance: success work* is a self assessed question on importance of success of work, ranging from 1 (very unimportant) to 4 (very important). *parents A-level*, *female*, and *IMC (fail)* are dummy variables if one of the parents had an A-level certificate, if the individual is female, and failed the Instructional Manipulation Check by Oppenheimer et al. (2009), respectively. *crtsuccess* denote the number of solved Cognitive Reflection Test questions (Frederick, 2005). *siblings* is the number of siblings. *grade A-level* is the self-reported A-level grade (*Ger. Abitur*). *politics left-right* is the self reported political orientation ranging from 1 (left) to 11 (right). Robust standard errors in parentheses. ***, **, and * denote significance with at the 1%, 5%, and 10% level, respectively.

Table A3.3: Comparison of standard errors

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $\widehat{WTPDIS}$ | 0.160*** | 0.160*** | 0.160*** | 0.160*** | 0.090** |
| | (0.037) | (0.044) | (0.024) | (0.024) | (0.038) |
| age | 0.032*** | 0.032*** | 0.032*** | 0.032*** | 0.032*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| female | -0.025 | -0.025 | -0.025** | -0.025** | -0.015 |
| | (0.016) | (0.019) | (0.012) | (0.011) | (0.012) |
| parents A-Level | 0.039*** | 0.039*** | 0.039*** | 0.039*** | 0.046*** |
| | (0.010) | (0.014) | (0.007) | (0.008) | (0.009) |
| openness | | | | | 0.004 |
| | | | | | (0.005) |
| extraversion | | | | | -0.013*** |
| | | | | | (0.005) |
| neuroticism | | | | | -0.040*** |
| | | | | | (0.005) |
| conscientiousness | | | | | -0.022*** |
| | | | | | (0.005) |
| agreeableness | | | | | -0.004 |
| | | | | | (0.006) |
| Constant | 1.546*** | 1.546*** | 1.546*** | 1.546*** | 1.550*** |
| | (0.058) | (0.063) | (0.046) | (0.047) | (0.045) |
| Standard errors | TS2SLS | Clustered | Bootstrap(50) | Bootstrap(500) | Clustered |
| Observations | 9,807 | 9,807 | 9,807 | 9,807 | 9,807 |
| Adj. $R^2$ | 0.069 | 0.069 | 0.069 | 0.069 | 0.077 |

*Notes:* This table reports different options for standard errors in the second stage of TS2SLS estimation. The first column reports the TS2SLS adjusted standard errors proposed by Pacini and Windmeijer (2016). Column (2) and (5) report standard errors cluster by individual. Column (3) and (4) reports the bootstrapped standard errors with 50 and 500 replications. The dependant variable is the natural logarithm of hourly labor earnings. $\widehat{WTPDIS}$ denotes the estimated willingness to pay in a disadvantageous situation. *parents A-level* and *female* are a dummy variables if one of the parents had an A-level certificate and if the individual is female. *extraversion*, *neuroticism*, *openness*, *conscientiousness*, and *agreeableness* denote the standardized Big-5 personality traits. Standard errors in parentheses. ***, **, and * denote significance with the highest standard errors at the 1%, 5%, and 10% level, respectively.

Do-File 3.1: TS2SLS adjusted standard errors

```
global y ln_hwage
global wlist age female puniabi
global wlist_num : list sizeof global(wlist)
global basicstd  di_impsuccess  bf_open_std- bf_agree_std
global basicstd_num : 5
/*Total number of instruments, here $basicstd, $wlist and const*/
global kz_num= ${basicstd_num} + ${wlist_num} + 1
global kx_num= 1 /*Number of predicted variables, here WTPDISADVC*/

reg WTPDISADVC ${basicstd} ${wlist}, r
reg WTPDISADVC di_impsuccess  ${wlist}
predict WTPDISADVChat_wobig5
reg ${y} ${basicstd}  ${wlist}  ${cond2}
est store eqn_y_redform
reg WTPDISADVC ${basicstd}  ${wlist}
est store eqn_x1
predict WTPDISADVChat ${cond2}

suest eqn_y_redform eqn_x1
/*Robust variance estimate of theta*/
/*without degrees of freedom correction*/
mat var = e(V)*(_N-1)/_N



/*Selecting rows and columns from var associated with theta*/
mata
/*Total number of instruments, here $basicstd, $wlist and const*/
kz = ${kz_num}
/*Number of variables in X, here WTPDISADVC, plus 1 for y*/
kyx = ${kx_num} + 1
```

```
sel = range(1,kz,1)
j=2
while (j<=kyx)
{
ss = range((j-1)*kz+j, j*kz+(j-1),1)
sel = sel\ss
j = j+1
}
var = st_matrix("var")
var = var[sel,sel]
st_matrix("Vthetahet",var)
end
drop if ${y}==.

/*Number of predicted variables, here WTPDISADVC*/
scalar kx = ${kx_num}
/*Number of exogenous variables, here wlist and constant*/
scalar ke = ${wlist_num} + 1
reg ${y} WTPDISADVChat ${wlist} ${cond2}
mat b2s = e(b)
mat b2sx = b2s[1,1..kx]'
local basicstd_num : list sizeof global(basicstd)
forval n=1/`basicstd_num'{
        local zn `: word `n' of ${basicstd}'
        qui regress `zn' WTPDISADVChat  ${wlist} ${cond2}

        if `n'==1{
        mat ch = e(b)'
        }
         else{
         mat ch = ch,e(b)'
         }
}
```

```
mat ch = ch,(J(kx,ke,0)\I(ke)
/* J(kx,ke,0) Description: the kx x ke matrix containing elements 0 */
/* I(ke) an ke x ke identity matrix*/

/*Calculating robust standard errors*/
mat delta = 1\-b2sx
mat var1het = (delta' # ch)*Vthetahet*(delta # ch')
mat seb2shet = vecdiag(cholesky(diag(vecdiag(var1het))))'

/*Displaying the results*/
local names = "WTPDISADVChat ${wlist} _cons"
mat colnames b2s = `names'
mat colnames var1het = `names'
mat colnames var1het = _:
mat rownames var1het = `names'
mat rownames var1het = _:
cap prog drop output2s
prog output2s, eclass
eret post b2s var1het
eret local depvar ${y}
eret local vcetype Robust
eret dis
end
output2s
```

Chapter 4

# Headscarf and job recruitment - Lifting the veil of Labour Market Discrimination

# 4 Headscarf and job recruitment - Lifting the veil of Labour Market Discrimination

*Co-authored with Attakrit Leckcivilize.*
*Published in IZA Journal of Labor Economics (https://doi.org/10.1186/s40172-018-0071-x)*
*The economics department of Leibniz Universität Hannover provided a funding of 3,520 EUR. This funding body did not influence the experimental design. This funding is generally rewarded to cover implementation costs of studies in experimental economics at the Leibniz Universität Hannover.*
*Written in British English.*

## 4.1 Introduction

The topic of discrimination based on, for example, ethnic origin and gender in the labour market came under scrutiny of the economics discipline after the influential doctoral dissertation entitled "The Economics of Discrimination" by the Nobel laureate Gary Becker in 1957. Becker proposed the concept of taste discrimination in which prejudiced persons receive disutility from their interaction with certain groups of people. Hence, they monetise their prejudice by applying a mark-up to the transaction. Even if two demographic groups were to have identical productive characteristics, such a mark-up leads to differences in compensation. In the labour market, taste discrimination can be classified by the source of the prejudice into employer, employee, and customer discrimination.

To consistently estimate the impacts of discrimination, researchers would ideally like to observe labour market performances of two groups which are the same in every aspect except for the characteristic of interest such as gender or ethnic origin (see Baert, 2018; Bertrand and Duflo, 2017; Lane, 2016; Zschirnt and Ruedin, 2016 for extensive review). Many studies (e.g. Bertrand and Mullainathan, 2004; Carlsson and Rooth, 2007) focused on racial discrimination at the very first stage of entry into the labour market and adopted the technique called "correspondence testing", which is to create fake CVs, allocate fake ethnicity at random to each CV and send these CVs to job adverts. They found that ethnic minorities received significantly fewer callbacks for job interviews. However, in some studies (e.g. Kaas and Manger, 2012), such a difference disappears when favourable information about the applicants' personality is included in the applications. They interpreted this finding as evidence for statistical discrimination[1].

In terms of appearance, Hamermesh and Biddle (1994) found significant effects of beauty on earnings in the US. Yet there is evidence of sorting by looks and beauty premium from some occupations, such as salespersons and lawyers, where the workers have to appear in public or confront the buyers directly (Biddle and Hamermesh, 1998). Hence, a fraction of the beauty premium could result from productive characteristics of beauty. A few studies used the correspondence testing technique in this domain as well. Rooth (2009) showed that an obese applicant received 20% fewer callbacks for an interview. Based on the correlation between job performance and being obese, he concluded that customer discrimination and/or statistical discrimination could be the explanation. On the other hand, Kraft (2012) found that unattractive candidates are 14% less likely to get an invitation and have to wait a couple of days longer for a callback. However, he did not find differential effects between high and low customer contact positions.

Recently, there are studies using correspondence testing to uncover discrimination against certain religious practices. In France, Valfort (2017) distinguished effects of applicants' religion from their country of origin by comparing the callback rates among fictitious candidates whose religion is Catholic, Judaism or Islam. All of them came from Lebanon, completed their high school, then received a certificate in Paris and became naturalised French citizens. She found that practising Catholic raises a callback rate by 30 and 100 percent respectively higher than being a Jew or being a Muslim. However, these disadvantages could be alleviated if the applicants signalled through extra-curricular activities in their CVs that they were secular rather than serious practitioners of these religions.

With regard to religious practice signalled by attire, Weichselbaumer (2016) sent out the same CV of a female candidate with different combination of names and photos to job openings for secretaries, accountants, and chief accountants in Germany. Particularly, she took photos of the same woman with and without headscarf and assigned either German or Turkish sounding names to the CV. Her results showed that a photo with German name was significantly more likely to get a callback than the same photo with Turkish name. In addition, a woman with a Turkish

[1] Statistical discrimination can arise when the recruiters face incomplete information about some unobservable characteristics of the applicants. Therefore, they have to form some beliefs about such characteristics of that applicant based on the "perceived" average of the groups (ethnicity, gender, or appearance) which the applicant belongs to (Altonji and Blank, 1999).

name and headscarf suffered additional discrimination compared to the same Turkish woman without headscarf.

Incorporating more variation in characteristics of both applicants and recruiters than previous studies, this paper demonstrates that interaction between religious practice and positive characteristics of applicants can mitigate impacts of discrimination, while it provides some evidence of additional heterogeneous effects based on recruiters' characteristics. Particularly, our study design combines a randomised CV approach with a laboratory experiment, which allows us to assess the effects of beauty, ethnicity, and religious practice in the same empirical model. We recruited students from universities in Hannover, Germany, to participate in an experiment where they were asked to select applicants for an interview for fictitious positions from a pool of candidates whose CVs were randomised in every other aspect except for appearance. An additional advantage of conducting such a lab experiment is to be able to control for personal characteristics of the participants (acting as HR recruiters) involved in the selection process[2].

We also tracked the time each participant used to evaluate each part of the presented CVs. This extra information allows us to (partially) control for the dual-process framework in judgement and decision making that could lead to bias in the hiring process (Derous et al., 2016). The dual-process theories involve Type 1 and Type 2 processes where the former is spontaneous, intuitive, effortless, and fast, while the latter is deliberate, rule-governed, effortful, and slow (Kahneman and Frederick, 2002). Specifically, we can control for the relative time each participant spends on each component of a particular CV, especially the photo page, with respect to his or her own average. Hence, any influences of heuristics involving Type 1 process could be partialled out from our results.

Exploiting a sizeable proportion of Turkish descendants in Germany, we randomly insert photos of the same Turkish looking women with and without headscarf into CVs in the experiment[3]. Apart from providing consistent estimates of the differences in the probability of being selected for an interview owing to beauty, ethnicity and headscarf, we attempt to identify the source of such discrimination based on job positions, characteristics of the CVs and the participants. Specifically, we classify our job openings into high and low skilled occupations as well as jobs

[2]There are very few studies based on the correspondence testing technique that can acquire information of HR personnel involved in the selection process. Rooth (2010) in Sweden is one paper that conducted follow-up surveys with the HR responsible for the recruitment of the positions, to whom the fake CVs were sent to.

[3]We decided not to include any German-looking woman with headscarf into this experiment because it is uncommon to see this group of population in everyday life. Otherwise, our focus on headscarf would become obvious to our participants.

with and without (or minimal) customer contact. We hypothesise recruiting for occupations with more interaction with customers would prefer better-looking persons and avoid minorities or females with headscarf due to anticipated customer discrimination, hence leading to potentially higher productivity such as higher sales. Conversely, any discrimination observed in low customer contact jobs such as back-office operations could mainly arise from either within firm (employer/employee) discrimination or statistical discrimination.

Our results suggest that the beauty premium prevails in all types of occupations and is quite large in high skilled occupations. So, the beauty premium could be driven by both taste discrimination and potential productive attributes of beauty. Nevertheless, a slightly larger premium in high skilled jobs supports the argument for employee discrimination because this is a sector where our participants could relate to the candidates as their future co-workers. Interestingly, better-looking candidates with the same gender as the recruiter are less likely to be chosen for the interview. Although a simple difference in the probability of being chosen between Turkish and German applicants shows a significant discrimination against Turks, such an effect disappears after controlling for beauty and interactions between some applicants' characteristics and headscarf. This finding provides an alternative explanation to the previous studies that racial discrimination in Germany might be partly explained by the fact that Turkish applicants (at least in our sample) are perceived as less beautiful than German looking counterparts.

Similar to Weichselbaumer (2016), Baert et al. (2017), and Valfort (2017), we find that positive characteristics mitigate negative impacts of religious practice and ethnicity in job recruitment. Such results provide circumstantial evidence for the importance of statistical discrimination (Haan et al., 2017) and biased beliefs (Reuben et al., 2014) as potential causes of discrimination against headscarf. In other words, supplying more productivity-relevant information could reduce average differences in perceived unobservable characteristics between candidates with and without headscarf. Yet unlike, for example, Baert and De Pauw (2014) who gather information on key attitudes underlying different mechanisms, our method is only an indirect way (yet popular among researchers) to try to isolate taste-based discrimination from statistical discrimination (Bertrand and Duflo, 2017). Specifically, the main goal of our study is to investigate when people discriminate against the headscarf, rather than why.

Despite the non-dynamic setting of our experiment, a reversal in discrimination effects against headscarf from over-penalised candidates with low characteristics to over-rewarded those with high characteristics is in line with a theory of dynamics of discrimination proposed by Bohren et al. (forthcoming). Using data from an online platform, they showed that women with no prior evaluation score on the platform were discriminated against but women with a history of positive evaluations were favoured over their male counterparts. Similar to our case, the recruiters may hold a certain kind of biased beliefs in abilities of women with a headscarf. Thus, extra information on positive characteristics helps to reverse negative effects of discrimination into advantages for this group of applicants.

Furthermore, the extent of discrimination against headscarf (conditioning on being Turkish) is more prominent in case of high skilled occupations and jobs with customer contact. Interestingly, we find evidence that such discrimination seems to be driven by male recruiters [4]. Older participants are those who discriminate against headscarf but at the same time they value good characteristics of these applicants. Our results imply that the characteristics and behaviours of the "recruiter" could be the main driver of observed discrimination, particularly against headscarf, during the hiring process. Yet such practices might not reflect the best interest of their employers in terms of firms' productivity or profit.

The paper is organised as follows. Section 4.2 explains the experimental design. Section 4.3 presents our methodology. Section 4.4 and 4.5 show the results and robustness checks, respectively. Section 4.6 discusses and concludes; Figures and Tables are included in the appendix.

## 4.2 Experimental design

The experiment is divided into two parts and all participants are students from local universities. Both parts were programmed using the software z-Tree (Fischbacher, 2007). The first part was carried out in December 2015 where 120 students had to act as HR personnel and choose candidates for a call back based on CVs. This part took one hour and the students were paid 20 Euro for participation. The decisions were not incentivised in order to capture an unbiased perception of the students. Otherwise, it might be possible that the participants would choose what they think is preferable by the researchers and do not reveal their true preferences. Descriptive statistics for the participants are summarized in Table 4.1. To assess the importance

[4] There are some indications from social psychology literature that men are more prone to express racial prejudice than women (Akrami et al., 2000).

of beauty on the probability of being selected, we conducted the second experiment, where the only task was to rate the photos from the first experiment. The rating was performed in March 2016, with 40 students in total. The second part took around 20 minutes and the students were paid 7 Euro. Both experiment took place in the computer lab of the Leibniz Universität Hannover in several sessions, each of them with 10 to 17 participants. We did not acquire any photos in Hannover, so the possibility that "recruiters" had ever met the candidates in real life is negligible.

As the objective of the first part was to simulate the recruiting process, participants were asked to select applicants for an interview for fictitious positions. For each position they selected applicants after reviewing seven characteristics, application photos[5] and the names of the applicant. The seven characteristics were presented in the following order:

- work experience (ranging from zero to three years)
- expected wage (average wage, 10% higher/lower than average, 20% higher/lower than average)
- Grade university/school (average, higher/lower than average)
- Quality of education: which is reputation of the college in the case of high skilled jobs or amount of absence days in the case of low skilled jobs[6] (average, higher/lower than average)
- Current unemployment spell (currently not unemployed, 1-6 months unemployed, 7-12 months unemployed, 13-18 months unemployed)
- Computer skills (sufficient, good, very good)
- English skills (basic, advanced, fluent)

The characteristics were randomly assigned, the photos appeared in a random order, and the names were randomly assigned (corresponding to gender and ethnic background of the photo). We presented the photos and the characteristics separately (see Figure A4.1 and Figure A4.2) so that we could record the time used in each part and analyse whether the amount of time participants looked at the photo is correlated with discrimination. As for headscarf, we constructed the control group by asking a professional photographer to take photos of three Turkish-looking

[5]It is common in Germany to send an application photo in a CV.

[6]The amount of absence days is reported on the school certificate and is an important criteria when hiring low skilled job entrants.

Table 4.1: Descriptive Statistics for the experiments

| | Mean | SD | Min | Percentile | | | Max |
|---|---|---|---|---|---|---|---|
| | | | | 25th | 50th | 75th | |
| **Panel A: Main experiment** | | | | | | | |
| female | 0.450 | 0.500 | | | | | |
| age | 23.208 | 3.149 | 18 | 21 | 23 | 24 | 45 |
| siblings | 1.364 | 0.975 | 0 | 1 | 1 | 2 | 4 |
| study semester | 4.065 | 2.675 | 1 | 2 | 3 | 6 | 15 |
| bachelor | 0.593 | 0.494 | | | | | |
| master | 0.278 | 0.450 | | | | | |
| born in Germany | 0.944 | 0.231 | | | | | |
| migration background | 0.185 | 0.390 | | | | | |
| **Panel B: Beauty rating** | | | | | | | |
| female | 0.525 | 0.506 | | | | | |
| age | 23.700 | 3.291 | 18 | 21 | 23 | 25 | 33 |
| siblings | 1.600 | 0.955 | 0 | 1 | 2 | 2 | 4 |
| study semester | 4.900 | 3.233 | 1 | 2 | 5 | 7 | 15 |
| bachelor | 0.575 | 0.501 | | | | | |
| master | 0.200 | 0.405 | | | | | |
| born in Germany | 0.900 | 0.304 | | | | | |
| migration background | 0.200 | 0.405 | | | | | |

*Notes:* Main experiment was conducted in December 2015, descriptive statistics for 109 participants are presented in Panel A. Beauty rating (Panel B) was conducted in March 2016 with 40 participants. SD = standard deviation

women with and without the headscarf, while keeping everything else the same (see Figure A4.3).

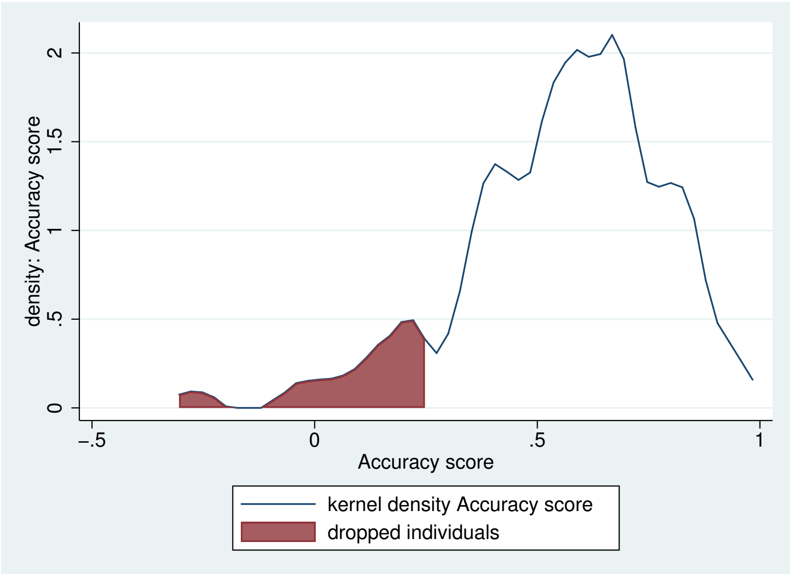
Each participant had to review 32-48 applications for 8-12 positions (see Table A4.1 and Table A4.2 for a full list of positions)[7]. These positions were organised in blocks of four positions. After each block the participants were asked to take a break and answer a questionnaire (paper-and-pencil questionnaire about socio-demographic characteristics) and then an experiment for another project dedicating to analyse methods of multidimensional scaling done by Jelnov (2018)[8]. We also used his data in our analysis in order to measure consistency of decision and

[7]We took the wording for tasks and requirements from actual job postings, but removing the complexity, so that it is easier to understand for our student participants.

[8]Jelnov's experiment consisted of two parts, where participants had to focus on similarity between fruits. In the first part the participants saw two pairs of fruits on the screen and they had to decide which of the pairs was more similar. Using a bubble sorting algorithm these tasks were repeated until the participants ordered each pair of fruits by similarity. In the second part the participants were presented with only one pair of fruits and were asked to

exclude 10% of participants with inconsistent decisions. We dropped the individuals based on the correlation coefficient between two part's of Jelnov's experiment (see Figure 4.1 for the distribution where an accuracy score of one indicates perfect consistency of decisions, while zero shows the answers are not correlated at all). We implicitly assumed that those who made inconsistent decisions in his part, which is unrelated to our research question, were more likely to make inconsistent decisions in our simplified recruiting process.

Figure 4.1: Distribution of Accuracy score and 10% dropped individuals




We assigned pictures, names, and characteristics to each CV in three steps. First, we restricted the pool of possible characteristics to 834 combinations, thereby we ensured that no applicant could have very high or very low scores in all dimension, i.e. that scoring low in one dimension increases the probability to score high in another dimension and vice versa. Hence, we eliminated almost all the chance that one CV could dominate another CV in all dimensions. Each CV was randomly assigned with one of these 834 combinations. In the second step, we assigned the photos for each CV so that in the first block (consisting of 16 CVs) three pictures of Turkish looking women would certainly appear but at random positions. In the second block where the applicants have to fill four more jobs (with another 16 applicants) three different photos of Turkish women would also be shown randomly. Our programme guaranteed that if a recruiter saw one Turkish looking women with the headscarf in the first block, she would evaluate the same Turkish women without the headscarf in the second block and vice versa. After random

rate "How similar are these two fruits" on a 5 point Likert scale. We measured consistency of these decisions by the correlation coefficient between these two parts. Then we excluded 10 percent of participants with the lowest consistency 'score' from our analysis. The results for the full sample are discussed in Section 4.5.

positions for three Turkish looking women were chosen, we would randomly assign photos for the other CVs out of a pool of 10 female and 13 male German looking photos, with the restriction that no photo could appear twice in one block.[9] In the last step we randomly assign first and last name to each photo according to the gender and ethnicity of the photo. So if the photo was from a German looking applicant, we would randomly assign one of the 50 most common German last names, whereas if the applicant is Turkish looking, we would assign one of 50 most common Turkish last names. Similarly we also had three pools of 50 most common first names for Turkish female, German female, and German male photos respectively in order to assign appropriate first names randomly.

We used pictures from 13 females and 13 males. For each of the three Turkish women we had two pictures, one with headscarf and one without. Therefore we used 29 different pictures in total. According to the actual number each photo appeared in our experiment, roughly 10% were photos from Turkish women with headscarf and roughly 10% were Turkish women without headscarf.[10] We ensured that all pictured persons were roughly in the same age, around 25 - 30 years old. We framed the task such that all presented candidates are in the begin of their career and satisfied the formal requirements for the respective position but differed by the seven characteristics only. The participants of the experiment were asked to carry out a pre-selection and choose their first and second preference out of four candidates for each position.

In order to investigate the potential productive characteristics of beauty (Biddle and Hamermesh, 1998), the jobs in each block can be classified into four groups by level of skill (high or low skilled) and interaction with customers (high or low levels of contact). Concerning the rating of characteristics of each photo, we ran part two of the experiment, where we asked another group of students to rate persons on the photos by five characteristics: beauty, trustworthiness, friendliness, intelligence and physical resilience. We standardised these rating scores within each rater, then averaged the scores for each photo across all raters. We use the double standardised beauty rating as a main explanatory variable.

[9] Some of the recruiters had to answer a third block with another four positions and 16 applicants as well. In this case the photos were inserted with an only restriction that it could not appear twice in the same block. The procedure of assigning the characteristics and names was identical to the first two blocks.

[10] Theoretically equal proportions of CVs between Turkish women with headscarf and without headscarf would have been desirable in terms of maximising statistical power, but in that case the objective of our research would become too obvious for our participants.

## 4.3 Methodology

We adopt the Linear Probability Model (LPM) to estimate the impacts of beauty, ethnicity, and headscarf on the chance of being selected for an interview by the participants in our experiment.

$$y_{ijk} = \beta_0 + X_i'\beta + Z_j'\gamma + B_k'\delta + Int_{ik}\theta + time_{ij} + D_i + \varepsilon_{ijk}$$

where $y_{ijk}$ is a dummy variable equal to 1 if CV $i$ with photo $k$ is chosen by "recruiter" $j$. $X_i'$ is a vector of the CV's seven characteristics discussed previously on page 56 while $Z_j'$ is a vector of "recruiter" $j$ characteristics from participant's responses to a questionnaire. $B_k'$ is a vector for our main explanatory variables (based on the photo $k$ attached to each CV), which are a female dummy if the applicant is a female, a composite beauty rating score of photo $k$, a dummy variable for ethnic Turkish, and a dummy if the applicant wears a headscarf.

$Int_{ik}$ are vectors of interaction terms between the photo's specific characteristic of interest, which is the headscarf, and selected CV's characteristics (subset of $X_i' \times B_k'$). These interactions should capture additional effects of having desirable characteristics among candidates wearing headscarf. Due to our moderate sample size of photos and participants, we decide not to code each CV characteristic as a categorical variable represented by a set of dummy variables. Instead, we redefine the levels of each characteristic as integers where the higher the value, the more preferable it is (0, 1, 2, 3, or 4 depending on the amount of levels in that characteristics). For instance, being unemployed for 13-18 months is coded as 0, while currently not unemployed is coded as 3. Further, we assess an effect of beauty rating when the candidate is the same gender as the recruiter (student participant) by adding an interaction term between beauty score and a dummy variable of whether the recruiter and the candidate are the same gender.

Since our variables of interest are drawn from the photo accompanying each CV, we control for proportion of time each participant $j$ looked at the photo page of CV $i$ relative to $j$'s total time used in that job position ($time_{ij}$). Meanwhile, $D_i$ are dummies for the order of CV $i$, i.e. applicant number 2, 3 or 4 in each job-position (with the first applicant as the reference group). These dummies are included to control for a tendency that some participants might systematically choose first, second, third or fourth applicant more often than other choices. Lastly, $\beta$, $\gamma$, $\delta$ and $\theta$ are vectors of parameters and $\varepsilon_{ijk}$ are the error terms which are clustered by 29 photos. As for robustness checks, we also cluster standard errors by participant as well as use two-way clustered robust standard errors by both photo and participant (Cameron and Miller, 2015).

In order to verify if the results are driven by characteristics of job openings or those of participants, we estimate the model for several subsamples based on job classifications (comprising high skilled, low skilled, with and without customer contact), gender of participants, age of participants (either older or younger than 23 years) and their total time used in the experiment.

We also explore the differential role of characteristics on the ranking chosen by participants in our experiment. Using the same regression specification as before, we redefine the dependent variable ($y_{ijk}$) from the first choice analysis to be a dummy variable for being chosen as the first preference, i.e. the second choice as well as those unselected are coded as 0. Regarding the second rank, the main LPM with $y_{ijk}$ equal to 1 for the second choice only is applied to a restricted sample dropping all the first choices. Of course, this method implicitly assumes that participants decided on their first preference candidate before comparing the remaining three candidates in order to select their second choice. We then estimate Conditional Logit model as a robustness check later in Section 4.5.

## 4.4 Results

Following the model in Section 4.3 we focus mainly on the results of the linear probability model. The results from participant fixed effects estimation are very similar to the OLS and are reported in Table 4.8. Table 4.2 shows that our randomisation process worked quite well and there is no significant difference in relevant characteristics between different sub-groups except for experience which is significantly different at the 10% level. Yet we always control for all CV's characteristics in our models. Table 4.3 shows simple regressions with dummy variables for gender, Turkish background, and headscarf. We do not observe significant discrimination against headscarf in this setting but only discrimination for Turkish background in high skilled jobs and jobs with less customer interaction. This result remains quite robust after including controls (see Table 4.4) for characteristics and the order a candidate appears in each job opening ($D_i$).

However, Table 4.5 shows that these negative and significant coefficients for Turkish looking applicants disappear after controlling for beauty, i.e. the lower chance of female candidates with

Table 4.2: t-test results by headscarf

| | Mean (no headscarf) | Mean (headscarf) | Difference | P-value |
|---|---|---|---|---|
| experience | 1.464 | 1.354 | 0.110 | 0.055 |
| wage | 2.115 | 2.034 | 0.081 | 0.238 |
| education | 0.989 | 1.029 | -0.039 | 0.360 |
| quality | 0.953 | 0.987 | -0.034 | 0.427 |
| unemployed | 1.566 | 1.581 | -0.015 | 0.791 |
| computer | 0.752 | 0.776 | -0.024 | 0.573 |
| english | 0.742 | 0.781 | -0.039 | 0.354 |
| Observations | 3544 | 384 | | |

*Notes:* Results for student's t test on difference of characteristics in the CVs by headscarf. The characteristics are rescaled so that a higher value is always better for the employer, see page 56 for a detailed discussion.

Turkish origin in our sample being selected results from their lower beauty rating comparing to German looking applicants[11]. This result is quite surprising because most studies on correspondence testing in different countries tend to find significant lower average callback rates for minorities. Yet there are some exceptions; for example, recent studies by Kraft (2012) in Germany and Edo et al. (2019) in France show no significant discrimination against female foreigners who signal good language skills. Owing to our focus on young applicants, all Turkish applicants are female and in their 20s. Hence, they correspond to second or third generation migrants, who in general speak German as their mother tongue. In addition, Table 4.4 and 4.5 show intuitive and robust results for the significance of the seven characteristics. We observe that labour market experience, final grade and quality of education increase the chances of being called back in every job category. Computer skill also has positive and significant effects but becomes less important among jobs with customer contact and low skilled jobs, whereas English skills increase the callback chances in occupations with customer interaction and high skilled jobs. Surprisingly, coefficients of previous unemployment and expected wage are not significant at the 5% level in any types of occupations. Table 4.5 also shows a very significant and robust effect of appearance on the chance of being selected. We do not observe a clear difference in the effect of beauty between jobs with and without customer interaction. The effect of beauty differs slightly between low skilled and high skilled occupations. Our experiment implies that the recruiters do not give extra rewards for potential productivity improvement of beauty in jobs

[11] There are roughly 20% of our participants in the main experiment and rating panel who have either non-German father or mother, so the migration background of participants should not drive such a finding.

Table 4.3: No controls, no beauty, and no interaction effects

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| Turkish origin | -0.067** | -0.106*** | -0.036 | 0.026 | -0.167*** |
| | (0.025) | (0.021) | (0.036) | (0.030) | (0.030) |
| headscarf | 0.035 | 0.116** | -0.035 | -0.026 | 0.101** |
| | (0.033) | (0.050) | (0.033) | (0.034) | (0.049) |
| photo female | 0.021 | -0.016 | 0.058** | 0.007 | 0.035 |
| | (0.021) | (0.026) | (0.026) | (0.025) | (0.026) |
| $2^{nd}$ position | 0.003 | 0.048 | -0.039 | -0.003 | 0.007 |
| | (0.019) | (0.035) | (0.034) | (0.026) | (0.036) |
| $3^{rd}$ position | 0.039 | 0.080** | -0.000 | 0.015 | 0.062* |
| | (0.025) | (0.037) | (0.034) | (0.034) | (0.031) |
| $4^{th}$ position | 0.046* | 0.111*** | -0.017 | 0.004 | 0.091** |
| | (0.023) | (0.039) | (0.033) | (0.023) | (0.036) |
| Observations | 3928 | 1960 | 1968 | 1944 | 1984 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

with customer interaction but favour beauty more in high skilled jobs. Perhaps, they can relate themselves to these high skilled occupations and prefer having more attractive people as their colleagues. The advantage of our laboratory experiment compared to a fake CV approach is our ability to control for characteristics of the recruiters and the time used for each part of the application. So we are able to detect that the beauty premium is only positive and significant for applicants of a different gender to the recruiter. This interaction effect significantly reduces the beauty premium for same gender applicants in Column (1) for the whole sample, and in Column (4) of the low skilled subsample.

Table 4.4: Controls, no beauty and interaction effects

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| Turkish origin | -0.063** | -0.085*** | -0.035 | 0.034 | -0.163*** |
| | (0.027) | (0.024) | (0.034) | (0.038) | (0.026) |
| headscarf | 0.027 | 0.088* | -0.037 | -0.033 | 0.102** |
| | (0.037) | (0.050) | (0.033) | (0.039) | (0.045) |
| photo female | 0.017 | -0.018 | 0.047* | -0.001 | 0.030 |
| | (0.019) | (0.023) | (0.025) | (0.024) | (0.024) |
| experience | 0.048*** | 0.033*** | 0.063*** | 0.054*** | 0.054*** |
| | (0.010) | (0.011) | (0.016) | (0.015) | (0.012) |
| wage | 0.008 | 0.009 | 0.007 | 0.022* | 0.003 |
| | (0.010) | (0.011) | (0.018) | (0.013) | (0.012) |
| education | 0.162*** | 0.144*** | 0.178*** | 0.140*** | 0.192*** |
| | (0.012) | (0.014) | (0.021) | (0.019) | (0.015) |
| quality | 0.116*** | 0.110*** | 0.125*** | 0.177*** | 0.063*** |
| | (0.014) | (0.015) | (0.020) | (0.017) | (0.014) |
| unemployed | 0.003 | -0.021 | 0.027* | 0.012 | 0.002 |
| | (0.009) | (0.014) | (0.014) | (0.014) | (0.013) |
| computer | 0.046*** | 0.060*** | 0.034* | 0.027* | 0.075*** |
| | (0.011) | (0.014) | (0.017) | (0.015) | (0.014) |
| english | 0.022* | -0.012 | 0.058*** | 0.016 | 0.035** |
| | (0.012) | (0.013) | (0.019) | (0.015) | (0.017) |
| time (photo) | 0.222* | 0.122 | 0.350* | 0.223 | 0.216 |
| | (0.128) | (0.179) | (0.180) | (0.172) | (0.155) |
| $2^{nd}$ position | 0.003 | 0.040 | -0.034 | -0.008 | 0.016 |
| | (0.018) | (0.033) | (0.030) | (0.025) | (0.032) |
| $3^{rd}$ position | 0.024 | 0.047 | -0.004 | -0.004 | 0.057* |
| | (0.024) | (0.034) | (0.031) | (0.034) | (0.029) |
| $4^{th}$ position | 0.039* | 0.096** | -0.018 | -0.004 | 0.084*** |
| | (0.021) | (0.036) | (0.030) | (0.024) | (0.029) |
| Observations | 3928 | 1960 | 1968 | 1944 | 1984 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

Table 4.5: Controls and beauty, no interaction effects

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| Turkish origin | 0.030 | 0.003 | 0.063 | 0.104* | -0.048 |
| | (0.035) | (0.038) | (0.050) | (0.055) | (0.034) |
| headscarf | 0.027 | 0.091** | -0.037 | -0.031 | 0.099** |
| | (0.030) | (0.044) | (0.026) | (0.033) | (0.039) |
| beauty | 0.064*** | 0.063*** | 0.063*** | 0.050** | 0.077*** |
| | (0.014) | (0.018) | (0.022) | (0.022) | (0.017) |
| photo female | -0.046** | -0.077*** | -0.019 | -0.050 | -0.044* |
| | (0.019) | (0.019) | (0.034) | (0.037) | (0.022) |
| quality | 0.116*** | 0.111*** | 0.125*** | 0.178*** | 0.062*** |
| | (0.014) | (0.015) | (0.020) | (0.017) | (0.014) |
| experience | 0.047*** | 0.032*** | 0.062*** | 0.054*** | 0.052*** |
| | (0.010) | (0.011) | (0.016) | (0.015) | (0.012) |
| education | 0.162*** | 0.144*** | 0.178*** | 0.141*** | 0.190*** |
| | (0.011) | (0.014) | (0.021) | (0.019) | (0.014) |
| english | 0.022* | -0.012 | 0.057*** | 0.017 | 0.033* |
| | (0.012) | (0.014) | (0.018) | (0.015) | (0.016) |
| unemployed | 0.003 | -0.020 | 0.027* | 0.013 | 0.002 |
| | (0.009) | (0.014) | (0.014) | (0.014) | (0.013) |
| computer | 0.046*** | 0.060*** | 0.033* | 0.027* | 0.076*** |
| | (0.011) | (0.014) | (0.017) | (0.015) | (0.014) |
| wage | 0.008 | 0.010 | 0.006 | 0.023* | 0.003 |
| | (0.010) | (0.010) | (0.018) | (0.013) | (0.012) |
| samegender | 0.022 | 0.029 | 0.014 | 0.004 | 0.037 |
| | (0.018) | (0.022) | (0.021) | (0.024) | (0.024) |
| samegender $\times$ beauty | -0.041** | -0.042* | -0.037 | -0.037* | -0.045 |
| | (0.020) | (0.024) | (0.025) | (0.021) | (0.028) |
| time (photo) | 0.221* | 0.131 | 0.343* | 0.229 | 0.203 |
| | (0.127) | (0.178) | (0.177) | (0.172) | (0.154) |
| $2^{nd}$ position | 0.002 | 0.036 | -0.033 | -0.009 | 0.015 |
| | (0.018) | (0.033) | (0.030) | (0.025) | (0.033) |
| $3^{rd}$ position | 0.022 | 0.043 | -0.005 | -0.004 | 0.051* |
| | (0.024) | (0.034) | (0.031) | (0.034) | (0.029) |
| $4^{th}$ position | 0.039* | 0.094** | -0.018 | -0.003 | 0.081*** |
| | (0.021) | (0.037) | (0.030) | (0.024) | (0.028) |
| Observations | 3928 | 1960 | 1968 | 1944 | 1984 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

Table 4.6: Controls and beauty, interaction effects

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| Turkish origin | 0.031 | 0.004 | 0.063 | 0.103* | -0.047 |
| | (0.035) | (0.038) | (0.050) | (0.055) | (0.034) |
| headscarf | -0.197*** | -0.200* | -0.206*** | -0.123* | -0.239*** |
| | (0.067) | (0.110) | (0.057) | (0.069) | (0.083) |
| headscarf × quality | 0.073** | 0.127*** | 0.022 | 0.039* | 0.118** |
| | (0.034) | (0.033) | (0.041) | (0.022) | (0.055) |
| headscarf × experience | 0.053*** | 0.046* | 0.060*** | 0.042** | 0.052*** |
| | (0.018) | (0.026) | (0.017) | (0.018) | (0.016) |
| headscarf × education | 0.039** | 0.082*** | 0.006 | -0.008 | 0.077** |
| | (0.019) | (0.025) | (0.025) | (0.027) | (0.035) |
| headscarf × english | 0.051*** | 0.023 | 0.076*** | 0.005 | 0.105*** |
| | (0.012) | (0.017) | (0.019) | (0.030) | (0.036) |
| beauty | 0.064*** | 0.063*** | 0.063*** | 0.050** | 0.077*** |
| | (0.014) | (0.018) | (0.022) | (0.022) | (0.017) |
| photo female | -0.046** | -0.079*** | -0.020 | -0.049 | -0.045** |
| | (0.019) | (0.020) | (0.034) | (0.037) | (0.022) |
| quality | 0.110*** | 0.100*** | 0.123*** | 0.175*** | 0.052*** |
| | (0.014) | (0.016) | (0.021) | (0.019) | (0.014) |
| experience | 0.043*** | 0.029** | 0.056*** | 0.051*** | 0.047*** |
| | (0.010) | (0.011) | (0.016) | (0.016) | (0.013) |
| education | 0.158*** | 0.140*** | 0.175*** | 0.142*** | 0.183*** |
| | (0.011) | (0.015) | (0.021) | (0.019) | (0.015) |
| english | 0.018 | -0.013 | 0.049** | 0.017 | 0.024 |
| | (0.013) | (0.015) | (0.018) | (0.017) | (0.017) |
| unemployed | 0.003 | -0.020 | 0.026* | 0.013 | 0.002 |
| | (0.009) | (0.014) | (0.013) | (0.014) | (0.013) |
| computer | 0.045*** | 0.061*** | 0.032* | 0.027* | 0.073*** |
| | (0.011) | (0.014) | (0.017) | (0.015) | (0.014) |
| wage | 0.008 | 0.010 | 0.006 | 0.023* | 0.002 |
| | (0.009) | (0.010) | (0.018) | (0.013) | (0.012) |
| samegender | 0.021 | 0.027 | 0.013 | 0.003 | 0.035 |
| | (0.017) | (0.021) | (0.020) | (0.024) | (0.023) |
| samegender × beauty | -0.041** | -0.040 | -0.036 | -0.037* | -0.044 |
| | (0.020) | (0.024) | (0.024) | (0.021) | (0.027) |
| time (photo) | 0.224* | 0.129 | 0.352* | 0.241 | 0.208 |
| | (0.128) | (0.177) | (0.178) | (0.174) | (0.153) |
| $2^{nd}$ position | 0.002 | 0.039 | -0.034 | -0.006 | 0.012 |
| | (0.018) | (0.032) | (0.030) | (0.025) | (0.032) |
| $3^{rd}$ position | 0.023 | 0.047 | -0.003 | -0.002 | 0.051* |
| | (0.024) | (0.034) | (0.031) | (0.035) | (0.029) |
| $4^{th}$ position | 0.040* | 0.096** | -0.017 | -0.002 | 0.082*** |
| | (0.021) | (0.036) | (0.030) | (0.024) | (0.029) |
| Observations | 3928 | 1960 | 1968 | 1944 | 1984 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

As for the headscarf, Table 4.6 shows that wearing headscarf decreases a probability of being selected in all types of occupations significantly for Turkish women (only significant at 10% level for low skilled and jobs with less customers contact). However, having good characteristics such as work experience and education also increase the chance of those with headscarf significantly. Such a mitigating role of favourable characteristics echoes findings from previous studies such as Valfort (2017), Kaas and Manger (2012), and Weichselbaumer (2016). Following the idea of belief-based discrimination proposed by Bohren et al. (forthcoming), we infer that our participants might have prior biases in beliefs about unobservable characteristics of candidates with headscarves. Yet positive signals from observable characteristics could reduce such an uncertainty and help the decision makers to rely less on their biased beliefs on the group average of these unobservables. Moreover, Bohren et al. (forthcoming) argued that knowing the existence of biased beliefs against themselves, those who are discriminated against have to put more effort and hence become better than the rest of the population despite the same observable characteristics. In our case, the decision makers might take this possibility into consideration. That is why we observe a reversal where good candidates with headscarves become more favourable than other good candidates without headscarves.

Our laboratory design also allows us to divide the sample by different subgroups. Table A4.3 shows that decision makers who take more time in our experiment favour less towards appearance. Although the magnitude of the headscarf-dummy is very high for fast participants, this coefficient is insignificant. Such a finding may result from the heterogeneity in this group, which should comprise some who did not take the experiment seriously (i.e. just randomly choose a candidate, hence should not discriminate) and others who did not carefully look at the characteristics but were guided by their intuition (heuristic judgement of Type 1 process) and thus discriminated more than the average. On the other hand, those who took more time for the experiment were more judicious (i.e. relying more on Type 2 process) and probably tried not to judge by appearance. Hence, the beauty premium is reduced for this subsample and only significant at 10% level. In Table A4.4, there is no significant difference between the genders in beauty premium but male participants seem to drive negative responses toward headscarf. As for age of participants we observe similar level of discrimination by appearance but the magnitude of discrimination against females and headscarf increases with the respondents' age.

Table 4.7: Controls and beauty, interaction effects

| Type of Occupation | All | | No Contact | | Contact | | Low Skill | | High Skill | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1st Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice |
| Turkish origin | 0.027 | 0.022 | -0.021 | 0.047 | 0.084** | 0.006 | 0.045 | 0.106 | 0.005 | -0.059 |
| | (0.026) | (0.041) | (0.022) | (0.048) | (0.038) | (0.058) | (0.048) | (0.080) | (0.039) | (0.042) |
| headscarf | -0.064 | -0.204*** | -0.155 | -0.149 | -0.014 | -0.247*** | 0.008 | -0.211** | -0.098 | -0.191*** |
| | (0.084) | (0.055) | (0.125) | (0.088) | (0.062) | (0.054) | (0.073) | (0.098) | (0.100) | (0.047) |
| headscarf × quality | 0.014 | 0.087** | 0.062*** | 0.123*** | -0.024 | 0.054 | -0.008 | 0.078*** | 0.027 | 0.117** |
| | (0.009) | (0.037) | (0.015) | (0.034) | (0.022) | (0.041) | (0.019) | (0.025) | (0.021) | (0.051) |
| headscarf × experience | 0.012 | 0.065*** | 0.023 | 0.049 | 0.011 | 0.066*** | -0.026 | 0.087*** | 0.044 | 0.034** |
| | (0.034) | (0.015) | (0.051) | (0.033) | (0.019) | (0.014) | (0.031) | (0.031) | (0.039) | (0.016) |
| headscarf × education | 0.006 | 0.044*** | 0.066** | 0.056* | -0.037* | 0.035 | -0.007 | 0.010 | 0.013 | 0.072* |
| | (0.020) | (0.014) | (0.026) | (0.033) | (0.019) | (0.021) | (0.030) | (0.027) | (0.024) | (0.037) |
| headscarf × english | 0.022 | 0.038 | 0.007 | 0.011 | 0.030 | 0.067** | -0.001 | -0.005 | 0.041 | 0.082** |
| | (0.016) | (0.023) | (0.022) | (0.017) | (0.018) | (0.032) | (0.032) | (0.021) | (0.039) | (0.039) |
| beauty | 0.035*** | 0.052*** | 0.013 | 0.072*** | 0.056*** | 0.034 | 0.022 | 0.046 | 0.047** | 0.058** |
| | (0.011) | (0.015) | (0.012) | (0.021) | (0.014) | (0.026) | (0.020) | (0.034) | (0.020) | (0.023) |
| photo female | -0.012 | -0.049* | -0.011 | -0.099*** | -0.019 | -0.004 | 0.022 | -0.086 | -0.048 | -0.013 |
| | (0.015) | (0.025) | (0.014) | (0.021) | (0.022) | (0.042) | (0.039) | (0.061) | (0.029) | (0.032) |
| quality | 0.090*** | 0.069*** | 0.096*** | 0.051*** | 0.084*** | 0.092*** | 0.140*** | 0.111*** | 0.044*** | 0.032** |
| | (0.009) | (0.015) | (0.015) | (0.017) | (0.014) | (0.024) | (0.014) | (0.020) | (0.011) | (0.015) |
| experience | 0.048*** | 0.014 | 0.044*** | 0.002 | 0.052*** | 0.029 | 0.055*** | 0.020 | 0.049*** | 0.018 |
| | (0.010) | (0.011) | (0.012) | (0.014) | (0.012) | (0.019) | (0.014) | (0.016) | (0.010) | (0.013) |
| education | 0.113*** | 0.111*** | 0.097*** | 0.106*** | 0.129*** | 0.121*** | 0.100*** | 0.104*** | 0.134*** | 0.126*** |
| | (0.008) | (0.014) | (0.012) | (0.016) | (0.011) | (0.025) | (0.015) | (0.021) | (0.010) | (0.018) |
| english | 0.015 | 0.010 | -0.008 | -0.011 | 0.039*** | 0.032 | 0.019 | 0.006 | 0.016 | 0.018 |
| | (0.010) | (0.013) | (0.016) | (0.017) | (0.011) | (0.020) | (0.014) | (0.018) | (0.013) | (0.016) |
| unemployed | 0.009 | -0.003 | -0.001 | -0.025* | 0.019 | 0.019 | 0.001 | 0.017 | 0.023* | -0.019 |
| | (0.010) | (0.007) | (0.015) | (0.013) | (0.013) | (0.016) | (0.012) | (0.014) | (0.013) | (0.012) |
| computer | 0.033*** | 0.031** | 0.041*** | 0.047*** | 0.029** | 0.015 | 0.019 | 0.019 | 0.053*** | 0.052*** |
| | (0.008) | (0.012) | (0.014) | (0.014) | (0.011) | (0.018) | (0.012) | (0.015) | (0.011) | (0.016) |
| wage | 0.005 | 0.007 | 0.011 | 0.008 | -0.000 | 0.008 | 0.018 | 0.015 | -0.001 | 0.004 |
| | (0.006) | (0.011) | (0.009) | (0.013) | (0.013) | (0.021) | (0.011) | (0.013) | (0.010) | (0.015) |
| samegender | 0.018 | 0.012 | 0.013 | 0.026 | 0.022 | -0.004 | 0.021 | -0.013 | 0.014 | 0.031 |
| | (0.013) | (0.018) | (0.016) | (0.024) | (0.020) | (0.020) | (0.016) | (0.028) | (0.021) | (0.024) |
| samegender × beauty | -0.020 | -0.034* | -0.004 | -0.045* | -0.031 | -0.021 | -0.028 | -0.026 | -0.012 | -0.038 |
| | (0.014) | (0.019) | (0.016) | (0.025) | (0.020) | (0.024) | (0.017) | (0.021) | (0.021) | (0.028) |
| time (photo) | 0.094 | 0.190 | 0.065 | 0.067 | 0.139 | 0.363 | 0.119 | 0.189 | 0.073 | 0.207 |
| | (0.118) | (0.149) | (0.160) | (0.154) | (0.162) | (0.234) | (0.176) | (0.191) | (0.143) | (0.187) |
| $2^{nd}$ position | 0.003 | 0.005 | 0.045* | 0.020 | -0.036 | -0.009 | -0.024 | 0.018 | 0.028 | -0.006 |
| | (0.018) | (0.022) | (0.024) | (0.038) | (0.029) | (0.027) | (0.021) | (0.030) | (0.030) | (0.032) |
| $3^{rd}$ position | 0.009 | 0.027 | 0.043* | 0.026 | -0.027 | 0.025 | -0.000 | 0.001 | 0.018 | 0.057 |
| | (0.017) | (0.027) | (0.023) | (0.040) | (0.027) | (0.033) | (0.029) | (0.037) | (0.023) | (0.037) |
| $4^{th}$ position | -0.079*** | 0.117*** | -0.050* | 0.162*** | -0.108*** | 0.066* | -0.105*** | 0.086*** | -0.057** | 0.151*** |
| | (0.016) | (0.024) | (0.026) | (0.038) | (0.025) | (0.034) | (0.024) | (0.024) | (0.026) | (0.035) |
| Observations | 3928 | 2946 | 1960 | 1470 | 1968 | 1476 | 1944 | 1458 | 1984 | 1488 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview as $1^{st}$ *Choice* or $2^{nd}$ *Choice* respectively. The number of observations for $2^{nd}$ *Choice* decreases because CV's which were selected as first preference are excluded from these estimations. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *migrback* is a dummy variables which is equal to 1 if the decision maker has a migration background. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

We present results from separated LPM for the first and second choice in Table 4.7. For the $1^{st}$ choice, the dependent variable is a dummy variable equal to 1 only if the candidate was selected as the first choice. In the columns with the $2^{nd}$ choice we drop those individuals who were selected as the first choice and perform another regression with a dummy dependent variable taking value 1 if the candidate was selected as the second choice or 0 otherwise. We observe that the headscarf mostly affects the second preference, i.e. when it is a borderline decision. Beauty also matters more among the second choice except for jobs with more customer contact, where beauty becomes more important for the first preference. We interpret this results as a sign that for the first preference the decision makers are more objective and focus on relevant characteristics such as work experiences, appearance, and gender for contact jobs. Yet for the second choice their decisions are nudged by other less important characteristics like religious practice, gender, or appearance in no contact jobs. In other words, the participants seem to discriminate in the situation where it is easier to discriminate because the decision is very close but they do not discriminate against the headscarf in cases where the woman with the headscarf is the most qualified out of the four, i.e. when discrimination is very costly to the employer.

## 4.5 Robustness analyses

We perform several robustness checks to assure that our findings are not sensitive to sample selection or model specifications. The full sample - including participants whose responses in another experiment were extremely inconsistent - is used to estimate regressions with four subgroups of job categories in Table A4.5[12]. The sign and significance of estimated coefficients are qualitatively similar albeit less precise. Using the main sample, we also estimate the Linear Probability Model with participant fixed effects and clustered standard errors at participant level. Controlling for unobservable heterogeneity across participants, LPM-FE should circumvent endogeneity concerns, i.e. non-zero correlation between CVs' characteristics such as Turkish background or wearing headscarf and participants' characteristics such as their migration background. Again, the results based on LPM-FE in Table 4.8 are very similar to those from LPM in Table 4.6. We also report two-way clustered robust standard errors by both photo and participant in Table A4.6.

[12]Results for other specifications are available upon request.

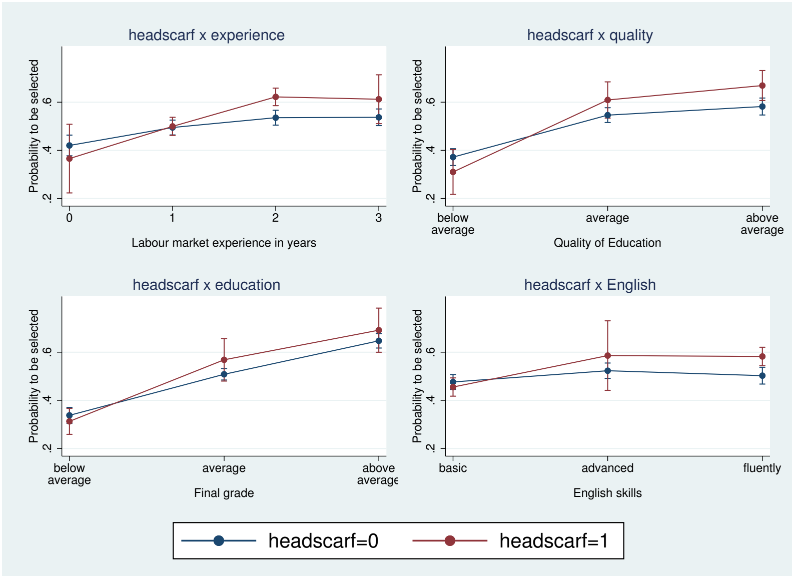
Table 4.8: Controls and beauty, interaction (by time), participant fixed effects

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| Turkish origin | 0.033 | 0.010 | 0.065 | 0.109* | -0.046 |
| | (0.040) | (0.057) | (0.058) | (0.065) | (0.055) |
| headscarf | -0.200** | -0.222* | -0.210 | -0.125 | -0.260** |
| | (0.088) | (0.131) | (0.132) | (0.130) | (0.125) |
| headscarf × quality | 0.076** | 0.137*** | 0.029 | 0.038 | 0.130*** |
| | (0.034) | (0.047) | (0.055) | (0.050) | (0.045) |
| headscarf × experience | 0.053** | 0.050 | 0.059* | 0.045 | 0.051 |
| | (0.023) | (0.038) | (0.034) | (0.035) | (0.033) |
| headscarf × education | 0.042 | 0.090* | 0.008 | -0.007 | 0.080* |
| | (0.035) | (0.046) | (0.047) | (0.054) | (0.047) |
| headscarf × english | 0.052 | 0.026 | 0.073 | 0.007 | 0.118** |
| | (0.034) | (0.050) | (0.050) | (0.053) | (0.052) |
| beauty | 0.067*** | 0.070*** | 0.069*** | 0.055** | 0.079*** |
| | (0.017) | (0.023) | (0.025) | (0.026) | (0.023) |
| photo female | -0.048* | -0.084** | -0.021 | -0.050 | -0.049 |
| | (0.027) | (0.036) | (0.040) | (0.042) | (0.038) |
| quality | 0.123*** | 0.122*** | 0.139*** | 0.196*** | 0.068*** |
| | (0.014) | (0.019) | (0.021) | (0.025) | (0.018) |
| experience | 0.053*** | 0.042** | 0.069*** | 0.064*** | 0.058*** |
| | (0.011) | (0.017) | (0.016) | (0.018) | (0.015) |
| education | 0.172*** | 0.161*** | 0.193*** | 0.163*** | 0.203*** |
| | (0.015) | (0.018) | (0.021) | (0.023) | (0.020) |
| english | 0.034** | 0.005 | 0.068*** | 0.034* | 0.046** |
| | (0.013) | (0.018) | (0.018) | (0.019) | (0.019) |
| unemployed | 0.010 | -0.012 | 0.035** | 0.021 | 0.009 |
| | (0.013) | (0.017) | (0.017) | (0.017) | (0.018) |
| computer | 0.063*** | 0.082*** | 0.050*** | 0.040** | 0.099*** |
| | (0.013) | (0.016) | (0.018) | (0.019) | (0.017) |
| wage | 0.019 | 0.023 | 0.016 | 0.034* | 0.013 |
| | (0.012) | (0.015) | (0.015) | (0.017) | (0.016) |
| samegender | 0.021 | 0.026 | 0.014 | 0.003 | 0.038* |
| | (0.017) | (0.024) | (0.024) | (0.026) | (0.023) |
| samegender × beauty | -0.044** | -0.048* | -0.042 | -0.043 | -0.042* |
| | (0.018) | (0.026) | (0.028) | (0.027) | (0.022) |
| time (photo) | 0.342*** | 0.310** | 0.461** | 0.392** | 0.389* |
| | (0.128) | (0.152) | (0.193) | (0.151) | (0.205) |
| $2^{nd}$ position | 0.001 | 0.037 | -0.034 | -0.009 | 0.011 |
| | (0.026) | (0.034) | (0.037) | (0.038) | (0.033) |
| $3^{rd}$ position | 0.022 | 0.043 | -0.005 | -0.004 | 0.050 |
| | (0.029) | (0.038) | (0.041) | (0.039) | (0.039) |
| $4^{th}$ position | 0.039* | 0.095*** | -0.017 | -0.003 | 0.081** |
| | (0.022) | (0.031) | (0.033) | (0.031) | (0.031) |
| Observations | 3928 | 1960 | 1968 | 1944 | 1984 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of participant are in parentheses. All regressions are estimated using LPM (Linear Probability Model). $^{***}$, $^{**}$ and $^{*}$ denote significance at the 1%, 5% and 10% level, respectively.

Since the main specification assumes that our seven characteristics in the CV can be coded as cardinal variables, we treat these characteristics as categorical variables represented by sets of dummy variables. Figure 4.2 illustrates the marginal effects of four main characteristics between candidates with and without headscarf on the predicted probability of being chosen. Although only some levels of these characteristics show significant differences in predicted probability of being chosen owing to the headscarf, their general trends conform to our main findings.

Figure 4.2: Predicted probability for a callback by categories of selected characteristics




As for the role of characteristics on ranking, instead of estimating two separated regressions for the first and second choices, we assume that our participants evaluate their first and second choices as a pair against all other possible pairs in each job opening. Hence, their response for each job opening can be rearranged from selecting 2 out of 4 candidates ($j$) to choosing 1 pair representing their first and second ranking out of all 12 possible pairs of the first and second choices ($k$). We then follow Cameron and Trivedi (2005) and specify the Conditional Logit Model with the probability of a pair of CVs $k$ being chosen by participant $j$ ($p_{jk}$) as follows:

$$p_{jk} = \frac{e^{\widetilde{\eta_{jk}}}}{\sum_{l=1}^{12} e^{\widetilde{\eta_{jl}}}}, k = 1, \ldots, 12$$

then

$$\begin{aligned}\widetilde{\eta_{jk}} = \widetilde{\beta}_0 + X'_{jk_1}\widetilde{\beta}^1 + X'_{jk_2}\widetilde{\beta}^2 + B'_{k_1}\widetilde{\delta}^1 + B'_{k_2}\widetilde{\delta}^2 \\ + Int_{ik_1}\widetilde{\theta}^1 + Int_{ik_2}\widetilde{\theta}^2 + D_{k_1} + D_{k_2} + time_{jk_1} + time_{jk_2}\end{aligned}$$

where $X'_{jk_s}$, $B'_{k_s}$, $Int_{ik_s}$, $D_{k_s}$ and $time_{jk_s}$ are vectors of the explanatory variables defined earlier with $k_s = k_1$ or $k_2$ standing for the first and second ranking in the pair $k$ respectively. Therefore, the main differences in explanatory variables between the main LPM and this Conditional Logit model are the inclusion of characteristics for both first and second choices and the exclusion of participant-level characteristics from the Conditional Logit model. Table 4.9 shows odds ratios of the chance that a pair of CVs is selected given a one unit change in particular characteristics of the first or second rank candidates with the Z-score in the squared brackets. Almost all of seven characteristics from both first and second CVs of a pair have odds ratios higher than one and are significant at the 1% level, i.e. the better the characteristics are, the more likely that pair of CVs is chosen. Odds ratios of beauty are also higher than one and significant at 5 or 10 % level in all job categories except for the beauty of the second rank in customer oriented jobs. Like our main results, beauty has negative effects on the chance of being selected if the decision maker and the candidate are of the same gender. Yet for the second choice such effects are significant only for jobs with less customer interaction and high skilled jobs at 5 and 10 % level, respectively.

In all jobs, wearing headscarf affects the chance of being selected as the second choice negatively but not for the first choice with significant and positive interaction effects for some characteristics of the CVs. In other words, the headscarf reduces the chances of being selected for applicants with unfavourable characteristics and increases the chances for applicants with good characteristics. Regarding subgroups, headscarf reduces the chance of being chosen (as both first and second choice) in high skilled jobs only.

One concern for the Conditional Logit Model is the validity of the independence of irrelevant alternatives (IIA) assumption. We test this assumption by comparing the estimated coefficients of the restricted model (excluding one choice combination) to the full model. The Wald test indicates that our model fails the IIA condition. This is plausible given that our choice combinations consist of, for example, a pair of $1^{st}$ and $2^{nd}$ CV, $1^{st}$ and $3^{rd}$ or $2^{st}$ and $3^{rd}$ CV. So it is hard to argue that the choice between any two of these combinations would be independent from other possible ‘irrelevant’ pair. Nonetheless, we decide to show these findings as a robustness check for our ‘two-step’ LPM.

Table 4.9: Conditional logit results (controls and beauty with interactions)

| Type of Occupation | All | | No Contact | | Contact | | Low Skill | | High Skill | |
|---|---|---|---|---|---|---|---|---|---|---|
| Turkish origin *rank1* | 1.166 | [0.80] | 0.912 | [-0.32] | 1.561 | [1.56] | 1.492 | [1.21] | 0.849 | [-0.54] |
| Turkish origin *rank2* | 1.097 | [0.47] | 1.115 | [0.42] | 1.060 | [0.20] | 1.624 | [1.58] | 0.729 | [-1.09] |
| headscarf *rank1* | 0.416 | [-1.52] | 0.156$^{*}$ | [-1.93] | 0.633 | [-0.56] | 0.907 | [-0.13] | 0.152$^{**}$ | [-2.03] |
| headscarf *rank2* | 0.322$^{**}$ | [-2.04] | 0.424 | [-1.03] | 0.257 | [-1.64] | 0.448 | [-1.09] | 0.197$^{**}$ | [-1.99] |
| headscarf *rank1* × quality *rank1* | 1.335 | [1.32] | 2.008$^{**}$ | [2.50] | 1.046 | [0.13] | 1.050 | [0.16] | 1.847$^{*}$ | [1.96] |
| headscarf *rank2* × quality *rank2* | 1.587$^{**}$ | [2.31] | 1.836$^{*}$ | [1.94] | 1.398 | [1.16] | 1.335 | [1.13] | 2.269$^{***}$ | [2.76] |
| headscarf *rank1* × experience *rank1* | 1.247 | [1.56] | 1.431 | [1.31] | 1.191 | [0.84] | 1.032 | [0.14] | 1.550$^{*}$ | [1.80] |
| headscarf *rank2* × experience *rank2* | 1.368$^{**}$ | [2.25] | 1.312 | [1.28] | 1.365 | [1.52] | 1.444$^{*}$ | [1.93] | 1.227 | [0.94] |
| headscarf *rank1* × english *rank1* | 1.247 | [1.22] | 1.175 | [0.54] | 1.263 | [0.89] | 0.978 | [-0.08] | 1.829$^{*}$ | [1.91] |
| headscarf *rank2* × english *rank2* | 1.291 | [1.33] | 1.197 | [0.69] | 1.411 | [1.19] | 1.025 | [0.10] | 1.837$^{**}$ | [2.16] |
| headscarf *rank1* × education *rank1* | 1.256 | [1.06] | 2.265$^{***}$ | [2.88] | 0.878 | [-0.42] | 0.965 | [-0.14] | 1.703$^{*}$ | [1.66] |
| headscarf *rank2* × education *rank2* | 1.348 | [1.35] | 1.467 | [1.25] | 1.256 | [0.79] | 1.035 | [0.11] | 1.770$^{**}$ | [1.97] |
| beauty *rank1* | 1.382$^{***}$ | [3.54] | 1.246$^{*}$ | [1.83] | 1.525$^{***}$ | [3.32] | 1.264$^{*}$ | [1.70] | 1.474$^{***}$ | [3.16] |
| beauty *rank2* | 1.316$^{***}$ | [3.44] | 1.447$^{***}$ | [3.46] | 1.198 | [1.47] | 1.285$^{**}$ | [2.02] | 1.365$^{***}$ | [2.63] |
| photo female *rank1* | 0.841 | [-1.27] | 0.793 | [-1.33] | 0.876 | [-0.64] | 0.970 | [-0.14] | 0.761 | [-1.42] |
| photo female *rank2* | 0.773$^{*}$ | [-1.91] | 0.627$^{***}$ | [-2.70] | 0.958 | [-0.21] | 0.666$^{**}$ | [-2.00] | 0.894 | [-0.60] |
| quality *rank1* | 2.837$^{***}$ | [9.46] | 2.952$^{***}$ | [7.23] | 2.842$^{***}$ | [6.78] | 3.510$^{***}$ | [7.48] | 1.992$^{***}$ | [5.20] |
| quality *rank2* | 1.772$^{***}$ | [5.83] | 1.583$^{***}$ | [4.07] | 2.021$^{***}$ | [5.11] | 1.939$^{***}$ | [5.03] | 1.503$^{***}$ | [3.26] |
| experience *rank1* | 1.718$^{***}$ | [6.97] | 1.625$^{***}$ | [4.01] | 1.847$^{***}$ | [5.70] | 1.690$^{***}$ | [4.29] | 1.824$^{***}$ | [6.08] |
| experience *rank2* | 1.239$^{***}$ | [2.92] | 1.132 | [1.45] | 1.352$^{***}$ | [2.77] | 1.196$^{*}$ | [1.95] | 1.310$^{***}$ | [2.96] |
| education *rank1* | 3.131$^{***}$ | [10.41] | 2.740$^{***}$ | [7.51] | 3.695$^{***}$ | [8.37] | 2.760$^{***}$ | [6.98] | 3.826$^{***}$ | [11.09] |
| education *rank2* | 2.055$^{***}$ | [7.99] | 1.936$^{***}$ | [6.64] | 2.208$^{***}$ | [5.88] | 1.854$^{***}$ | [4.96] | 2.372$^{***}$ | [7.74] |
| english *rank1* | 1.475$^{***}$ | [4.60] | 1.180 | [1.34] | 1.838$^{***}$ | [4.90] | 1.360$^{**}$ | [2.34] | 1.612$^{***}$ | [4.24] |
| english *rank2* | 1.237$^{**}$ | [2.33] | 1.073 | [0.68] | 1.400$^{***}$ | [2.60] | 1.109 | [0.82] | 1.382$^{***}$ | [3.01] |
| unemployed *rank1* | 1.239$^{***}$ | [2.68] | 1.112 | [0.97] | 1.413$^{***}$ | [3.07] | 1.181 | [1.49] | 1.343$^{***}$ | [3.13] |
| unemployed *rank2* | 1.089 | [1.11] | 0.958 | [-0.49] | 1.233$^{*}$ | [1.93] | 1.156 | [1.60] | 1.039 | [0.43] |
| computer *rank1* | 1.735$^{***}$ | [6.17] | 1.790$^{***}$ | [4.88] | 1.716$^{***}$ | [4.23] | 1.442$^{***}$ | [2.78] | 2.201$^{***}$ | [7.06] |
| computer *rank2* | 1.396$^{***}$ | [4.32] | 1.452$^{***}$ | [4.09] | 1.311$^{**}$ | [2.29] | 1.222$^{*}$ | [1.89] | 1.659$^{***}$ | [4.78] |
| wage *rank1* | 1.268$^{***}$ | [2.91] | 1.293$^{**}$ | [2.21] | 1.260$^{**}$ | [2.24] | 1.338$^{**}$ | [2.34] | 1.228$^{**}$ | [2.24] |
| wage *rank2* | 1.163$^{**}$ | [2.05] | 1.130 | [1.57] | 1.180 | [1.55] | 1.159 | [1.54] | 1.163$^{*}$ | [1.73] |
| samegender *rank1* | 1.133 | [1.34] | 1.079 | [0.64] | 1.157 | [1.20] | 1.107 | [0.72] | 1.192 | [1.46] |
| samegender *rank2* | 1.061 | [0.73] | 1.104 | [0.81] | 1.011 | [0.10] | 0.948 | [-0.49] | 1.205 | [1.61] |
| samegender *rank1* × beauty *rank1* | 0.826$^{**}$ | [-2.21] | 0.872 | [-1.00] | 0.808 | [-1.53] | 0.814 | [-1.53] | 0.862 | [-1.13] |
| samegender *rank2* × beauty *rank2* | 0.834$^{**}$ | [-2.08] | 0.768$^{**}$ | [-2.12] | 0.905 | [-0.77] | 0.867 | [-1.06] | 0.805$^{*}$ | [-1.72] |
| time (photo) *rank1* | 7.420$^{**}$ | [2.24] | 10.556$^{**}$ | [1.99] | 7.592 | [1.56] | 6.297 | [1.54] | 8.117 | [1.48] |
| time (photo) *rank2* | 9.196$^{**}$ | [1.98] | 5.199 | [1.27] | 18.507$^{*}$ | [1.79] | 5.835 | [1.27] | 12.440$^{*}$ | [1.88] |
| Choice combinations | 11784 | | 5880 | | 5904 | | 5832 | | 5952 | |

*Notes:* This table shows the odds ratios for the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. Odds ratios higher than 1 indicate higher chances for a CV to be selected, odds ratios less than 1 indicate a lower probability, and values of 1 indicate no change in probabilities. Z-scores are reported in squared brackets. Choice combinations denote the number of possible choices, for each decision the decision maker have 12 possible choices (instead of 4 compared to LPM), therefore the nuber increases by the factor of 3. $^{***}$, $^{**}$, and $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

## 4.6 Discussion & Conclusion

Exploiting a German practice to include photos in a CV allows us to measure how differences in appearance can influence job recruitment. Our laboratory experiment contributes to the sizeable literature on correspondence testing (fake CVs) in many aspects. Our research design enables us to gather information on socio-economic background of our participants as well as how long they look at the photo of each applicant. We also consider different aspects of appearance, i.e. beauty, ethnicity, and religious attire simultaneously. Further, based on job categories our paper can shed light on the sources of discrimination. Since the participants were asked to fill fictitious positions, the sources of discrimination against appearance could come from their taste based discrimination, statistical discrimination or productive characteristics of appearance. In correspondence testing experiments, the composition of a team might be an additional source of discrimination owing to HR's endeavour to hire someone who fit in with the existing team. However, our participants did not have such information; hence, we can rule out this channel.

Our findings suggest significant beauty effects on the hiring decision in every job category and such effects do not depend on gender or age of the recruiters. Although we do not find any discrimination against Turkish applicants after controlling for beauty, our results indicate a significant discrimination against those Turkish looking candidates who wear headscarf. Yet desirable characteristics of these applicants do help to reduce or even reverse their disadvantages. The older subgroup seems to discriminate more against headscarf but the decision is not driven by how long they look at the photos[13]. Looking closely into the mechanisms, employer discrimination might be underestimated in our experiment because the participants are young and have little labour market experience. Regarding customer discrimination, the negative and significant effects in more customer oriented occupations signal our participants' concern about perception of customers on headscarf. However, the negative result is not robust to an inclusion of participant-level fixed effects; thus, it could just be a biased estimate due to some unobservable characteristics of the participants.

[13]Such a finding does not provide strong evidence supporting ideas of dual-process framework or attention discrimination (where decision makers disfavour acquiring information about individuals from a priori less attractive group when asked to select only top applicants from a pool of candidates (Bartoš et al., 2016)). If there still exist heuristic judgement or attention discrimination against headscarf in our experiment, we already control for them through an inclusion of relative time used.

The only subgroup with a robust negative result is the high skilled. Since it is the only job category in which our participants can imagine themselves working with the fictitious candidates in the future, we interpret this as supporting evidence for employee discrimination. Also, compared to low skilled occupations, the high skilled could be perceived as requiring more personal interaction with other colleagues or managers. Furthermore, good characteristics such as better education and more work experience are being extra rewarded for candidates with headscarf. This implies that statistical discrimination could play an important role during the selection process. On balance, our participants might feel that they are not particularly against headscarf because they favour those with good characteristics, while being harder on candidates with headscarf and undesirable characteristics.

Unfortunately, the choice between correspondence testing and laboratory experiment comes with trade-offs. Despite being able to control for more variables, there are several reasons why our results might not reflect the true level of discrimination in the labour market. First, the participants were instructed by the experimenter and knew that they were observed. Therefore, the "true" effect might be shrouded by experimenter demand effects Zizzo (2010), where people tend to act according to social norms albeit contradicting their true beliefs. This could result in a much lower magnitude of discrimination against the headscarf compared to a field experiment (e.g. Weichselbaumer, 2016). Additionally, our experimental design fixes number of applicants to four per job opening and the participants were asked to always choose two of them. This is of course an unrealistic restriction in real world recruitment. Moreover, according to Becker (1957), discrimination should be more prevalent in industries with highly competitive labour markets because it is cheaper for firms to discriminate. Our setting, however, does not consider such a competition effect.

Despite our effort to mimic the field as close as possible, we are aware that our experiment setting most likely underestimates the true level of taste based discrimination in the labour market, hence jeopardising the external validity. Nevertheless, our experiment sheds some light on the mechanism behind discrimination, i.e. the participants discriminated against minorities with undesirable characteristics, but balanced this out by favouring those with good characteristics. The experiment suggests that good characteristics may not only compensate for but reverse the biased negative perception of minorities in the labour market through possibly implicit or subconscious affirmative action.

Our findings are in line with the model for dynamic discrimination by Bohren et al. (forthcoming), who show that a reversal of discrimination can occur if some evaluators hold a biased stereotype against a certain group, while others are aware of this. The basic idea is that at the first stage where the quality of the applicant is unknown a certain group is discriminated against due to biased beliefs, but the evaluators account for that at later stages after they received a positive signal of applicant's quality. In our experiment this is signalled by good characteristics (e.g. labour market experience or good educational outcomes). Interestingly, there are also studies which show that racial discrimination in the USA is higher if a signal of high productivity was included (Nunley et al., 2015). Our results indicate the opposite, namely that the discrimination is reduced if a signal for high expected productivity e.g. good grades and labour market experience is included. This difference might come from the fact that many field experiments use the first names as a signal for ethnicity. If names carry additional information about the applicant such as socioeconomic status (Fryer Jr and Levitt, 2004), this might bias the results, especially when the socioeconomic status is more important for high productivity applicants[14].

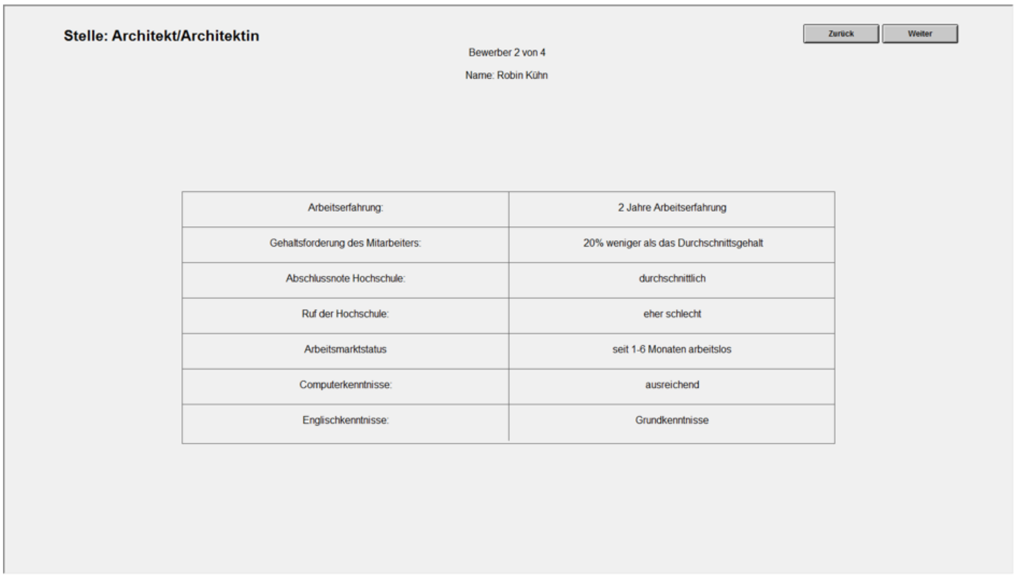
As for policy implication, one might wonder if a policy prohibiting a photo from CV could help reduce discrimination against appearance. In France Manant et al. (2019) found that despite having no photo on the CV, recruiters did gather information about their fictitious candidates' looks and religious practice from their Facebook profiles. Although photos make ethnicity of candidates clear to the recruiters, such disadvantages can be mitigated if the candidates from ethnic minorities look attractive, friendly or likable (Weichselbaumer and Schuster, 2017). Such a finding is consistent with our robust beauty premium. Further, our results suggest that the gap in job opportunities due to religious practice like headscarf could be narrowed down through a signalling of preferable characteristics. This gives room for policy interventions such as education or apprenticeship programmes targeting these groups. We hope that our paper will spur more discussion and encourage future research to consider appearance as a package of, for instance, beauty, ethnicity and observable religious practice simultaneously.

[14]Nunley et al. (2015) use only 8 names, namely DeShawn, DeAnde, Ebony, and Aaliyah as "black" names and Amy, Claire, Cody, and Jake that the applicant is "white". They try to address this criticism by adding a signal for socioeconomic status, namely a street address with high or low house prices . This approach will certainly reduce the bias, but it remains uncertain whether it is completely eliminated, i.e. whether all the recruiters know the street names. Additionally, Gebauer et al. (2012) find that popular names are a very significant predictor for clicks on a German dating site and also attractiveness of names to be highly correlated with self-esteem, smoking behaviour, and educational outcomes. The concern that black sounding names might carry more information than just the signal for ethnicity is also supported by the results of Fryer Jr and Levitt (2004), who show that the raw differences in various socioeconomic outcomes between black sounding and white sounding names are largely reduced or vanish completely if control variables for the child's circumstances at birth are added.

## 4.7 Appendix

Figure A4.1: Example of the experiment screen 1 (the photo is anonymised here for privacy reasons)


Figure A4.2: Example of the experiment screen 2




| | |
|---|---|
| Arbeitserfahrung: | 2 Jahre Arbeitserfahrung |
| Gehaltsforderung des Mitarbeiters: | 20% weniger als das Durchschnittsgehalt |
| Abschlussnote Hochschule: | durchschnittlich |
| Ruf der Hochschule: | eher schlecht |
| Arbeitsmarktstatus | seit 1-6 Monaten arbeitslos |
| Computerkenntnisse: | ausreichend |
| Englischkenntnisse: | Grundkenntnisse |

Figure A4.3: An example of photos of the same person with and without veil (the photos are anonymised here for privacy reasons)

Table A4.1: Job description (low-skilled jobs)

| **Job** | **Tasks** | **Requirements** |
|---|---|---|
| Tram driver | • Responsible tram driving<br>• Ensure road and operational safety<br>• Immediate response to fault and emergency | • Reliability and sense of responsibility<br>• Physical Resilience |
| Baker | • Preparing and making bakery<br>• Keeping equipment hygienic and clean<br>• Quality inspections and management | • Reliability and sense of responsibility<br>• Physical Resilience |
| Car Painter | • Apply primers and repaint car bodies<br>• Sand the base coat, paint, and polish small damages<br>• Check for runs in the paint to ensure the quality | • Reliability, sense of responsibility and ability to work autonomously<br>• Shift duty |
| Hotel manager | • Coordinating with cleaning and kitchen staffs<br>• Serve in the restaurant<br>• bookkeeping and inventory holding | • Communication skills<br>• Ability to work in a team and responsibility |
| Trained retail salesperson | • Consulting and Sales<br>• Inventory and cash management<br>• Quality inspection of goods | • Communication skills<br>• Responsibility and good teamwork |
| Childcare worker | • Organise activities and implement a curriculum<br>• Close cooperation with the parents<br>• Create a warm atmosphere with structured routines | • Ability to work in a team and communicate with parents<br>• Constructive implementation of our educational concept |

Table A4.2: Job description (high-skilled jobs)

| Job | Tasks | Requirements |
|---|---|---|
| Quantitative Risk Management Analyst | • Valuation of financial instruments<br>• Risk simulation and stress tests<br>• Model the market risk | • Ability to work independently and autonomously<br>• Structured documentation of your work progress |
| Auditor | • Preparing and checking financial reports<br>• Excellent knowledge of finance and tax laws<br>• Finding approaches to deal with complicated taxation issues | • Ability to work independently and autonomously<br>• Structured documentation of your work progress |
| Electronics Engineer | • Reliable evaluation of electronic systems<br>• Document compliance with safety standards<br>• Designing and installing of electric components in different projects | • Systematic way of thinking and organisational skills<br>• Enthusiasm for efficient and structured developing |
| Architect | • Pre/post tender discussions and customer service<br>• Design of detached houses<br>• Gaining new customers after being promoted | • Self-confident manner<br>• Ability to work in a team and communication skills |
| Civil Rights Attorney | • Providing legal advise to customers in all civil rights issues<br>• Negotiating settlements of legal disputes<br>• Gaining new customers after being promoted | • Self-confident manner<br>• Ability to work in a team and communication skills |
| Field veterinarian | • Customer liaison and support<br>• Building business relations<br>• Autonomously plan and coordinate on case-by-case services | • Customer-oriented work in an ambitious team<br>• Well-organised and target-oriented personality with excellent communication skills |

Table A4.3: Controls and beauty, interaction (by time)

| Decision maker | All | Fast | Average | Slow |
|---|---|---|---|---|
| Turkish origin | 0.031 | 0.028 | 0.059 | 0.019 |
| | (0.035) | (0.076) | (0.048) | (0.036) |
| headscarf | -0.197*** | -0.319 | -0.177 | -0.066 |
| | (0.067) | (0.190) | (0.104) | (0.240) |
| headscarf × quality | 0.073** | 0.103** | 0.122 | -0.004 |
| | (0.034) | (0.050) | (0.074) | (0.081) |
| headscarf × experience | 0.053*** | 0.062 | 0.028 | 0.067 |
| | (0.018) | (0.041) | (0.029) | (0.054) |
| headscarf × education | 0.039** | 0.133*** | 0.001 | 0.000 |
| | (0.019) | (0.034) | (0.020) | (0.042) |
| headscarf × english | 0.051*** | 0.020 | 0.056 | 0.033 |
| | (0.012) | (0.096) | (0.060) | (0.082) |
| beauty | 0.064*** | 0.076** | 0.085*** | 0.033* |
| | (0.014) | (0.032) | (0.024) | (0.017) |
| photo female | -0.046** | -0.061 | -0.041 | -0.032 |
| | (0.019) | (0.042) | (0.036) | (0.029) |
| quality | 0.110*** | 0.096*** | 0.115*** | 0.118*** |
| | (0.014) | (0.026) | (0.021) | (0.025) |
| experience | 0.043*** | 0.011 | 0.051*** | 0.065*** |
| | (0.010) | (0.021) | (0.016) | (0.016) |
| education | 0.158*** | 0.155*** | 0.142*** | 0.175*** |
| | (0.011) | (0.024) | (0.020) | (0.015) |
| english | 0.018 | 0.002 | 0.026 | 0.020 |
| | (0.013) | (0.022) | (0.023) | (0.016) |
| unemployed | 0.003 | -0.015 | 0.010 | 0.015 |
| | (0.009) | (0.021) | (0.013) | (0.015) |
| computer | 0.045*** | 0.051** | 0.021 | 0.062*** |
| | (0.011) | (0.019) | (0.019) | (0.019) |
| wage | 0.008 | -0.002 | 0.012 | 0.014 |
| | (0.009) | (0.021) | (0.016) | (0.015) |
| samegender | 0.021 | 0.057* | -0.000 | 0.003 |
| | (0.017) | (0.031) | (0.030) | (0.018) |
| samegender × beauty | -0.041** | -0.061* | -0.039 | -0.024 |
| | (0.020) | (0.032) | (0.038) | (0.018) |
| time (photo) | 0.224* | 0.010 | 0.133 | 0.454** |
| | (0.128) | (0.211) | (0.246) | (0.213) |
| $2^{nd}$ position | 0.002 | -0.030 | 0.025 | 0.012 |
| | (0.018) | (0.032) | (0.029) | (0.039) |
| $3^{rd}$ position | 0.023 | 0.038 | 0.002 | 0.033 |
| | (0.024) | (0.043) | (0.041) | (0.038) |
| $4^{th}$ position | 0.040* | 0.026 | 0.046 | 0.049 |
| | (0.021) | (0.036) | (0.034) | (0.037) |
| Observations | 3928 | 1288 | 1288 | 1352 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

Table A4.4: Controls and beauty, interaction (by subgroups)

| Decision maker | All | Male | Female | Age≤22 | Age≥23 |
|---|---|---|---|---|---|
| Turkish origin | 0.031 | -0.008 | 0.087 | 0.015 | 0.044 |
| | (0.035) | (0.037) | (0.083) | (0.065) | (0.035) |
| headscarf | -0.189** | -0.220*** | -0.127 | -0.100 | -0.256** |
| | (0.072) | (0.078) | (0.090) | (0.142) | (0.105) |
| headscarf × quality | 0.078** | 0.093** | 0.054 | 0.084 | 0.068* |
| | (0.035) | (0.044) | (0.038) | (0.091) | (0.034) |
| headscarf × experience | 0.054*** | 0.054** | 0.051** | 0.026 | 0.076** |
| | (0.018) | (0.024) | (0.020) | (0.017) | (0.027) |
| headscarf × education | 0.042* | 0.074*** | 0.006 | 0.075 | 0.024 |
| | (0.021) | (0.014) | (0.041) | (0.061) | (0.035) |
| headscarf × english | 0.054*** | 0.025 | 0.066** | 0.021 | 0.076*** |
| | (0.013) | (0.019) | (0.028) | (0.039) | (0.022) |
| headscarf × migrback | -0.099 | -0.191*** | -0.039 | -0.152 | -0.067 |
| | (0.061) | (0.048) | (0.078) | (0.131) | (0.089) |
| beauty | 0.064*** | 0.051** | 0.046*** | 0.069** | 0.058*** |
| | (0.014) | (0.021) | (0.016) | (0.028) | (0.019) |
| photo female | -0.046** | -0.049** | -0.048 | -0.039 | -0.059** |
| | (0.019) | (0.023) | (0.040) | (0.046) | (0.028) |
| quality | 0.110*** | 0.110*** | 0.109*** | 0.110*** | 0.109*** |
| | (0.014) | (0.018) | (0.024) | (0.020) | (0.019) |
| experience | 0.043*** | 0.031** | 0.057*** | 0.048** | 0.039*** |
| | (0.010) | (0.015) | (0.017) | (0.018) | (0.013) |
| education | 0.158*** | 0.165*** | 0.152*** | 0.166*** | 0.150*** |
| | (0.011) | (0.016) | (0.019) | (0.016) | (0.015) |
| english | 0.018 | 0.015 | 0.024 | 0.032* | 0.007 |
| | (0.013) | (0.016) | (0.021) | (0.018) | (0.018) |
| unemployed | 0.004 | 0.011 | -0.007 | 0.030* | -0.017 |
| | (0.009) | (0.013) | (0.016) | (0.016) | (0.013) |
| computer | 0.045*** | 0.042*** | 0.049** | 0.057*** | 0.035** |
| | (0.011) | (0.015) | (0.018) | (0.014) | (0.016) |
| wage | 0.008 | 0.011 | 0.005 | 0.028* | -0.007 |
| | (0.009) | (0.014) | (0.016) | (0.015) | (0.011) |
| samegender | 0.022 | | -0.006 | 0.045 | |
| | (0.018) | | (0.026) | (0.026) | |
| samegender × beauty | -0.041** | -0.013 | -0.005 | -0.045 | -0.035 |
| | (0.020) | (0.033) | (0.039) | (0.028) | (0.029) |
| time (photo) | 0.234* | 0.155 | 0.338* | 0.123 | 0.357** |
| | (0.128) | (0.186) | (0.183) | (0.231) | (0.144) |
| $2^{nd}$ position | 0.001 | 0.016 | -0.020 | 0.024 | -0.019 |
| | (0.018) | (0.027) | (0.030) | (0.030) | (0.027) |
| $3^{rd}$ position | 0.023 | 0.078** | -0.045 | 0.023 | 0.021 |
| | (0.025) | (0.030) | (0.033) | (0.032) | (0.028) |
| $4^{th}$ position | 0.039* | 0.049 | 0.022 | 0.056* | 0.025 |
| | (0.021) | (0.029) | (0.030) | (0.031) | (0.028) |
| Observations | 3928 | 2168 | 1760 | 1712 | 2216 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *migrback* is a dummy variables which is equal to 1 if the decision maker has a migration background. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

Table A4.5: Controls and beauty, interaction effects (full sample)

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| Turkish origin | 0.051 | 0.017 | 0.090* | 0.101* | -0.012 |
| | (0.033) | (0.040) | (0.047) | (0.051) | (0.031) |
| headscarf | -0.183** | -0.215* | -0.168** | -0.113* | -0.223** |
| | (0.074) | (0.115) | (0.073) | (0.066) | (0.098) |
| headscarf × quality | 0.060 | 0.118*** | 0.003 | 0.029 | 0.100* |
| | (0.038) | (0.037) | (0.038) | (0.023) | (0.058) |
| headscarf × experience | 0.047*** | 0.057** | 0.037* | 0.046*** | 0.037* |
| | (0.011) | (0.023) | (0.019) | (0.016) | (0.019) |
| headscarf × education | 0.042** | 0.076*** | 0.020 | -0.009 | 0.081** |
| | (0.021) | (0.025) | (0.020) | (0.025) | (0.035) |
| headscarf × english | 0.038** | 0.024 | 0.050 | -0.008 | 0.092* |
| | (0.018) | (0.019) | (0.034) | (0.021) | (0.045) |
| beauty | 0.064*** | 0.063*** | 0.066*** | 0.043** | 0.081*** |
| | (0.013) | (0.016) | (0.021) | (0.019) | (0.016) |
| photo female | -0.053*** | -0.083*** | -0.027 | -0.054* | -0.048** |
| | (0.016) | (0.021) | (0.030) | (0.031) | (0.021) |
| quality | 0.110*** | 0.102*** | 0.120*** | 0.170*** | 0.055*** |
| | (0.014) | (0.013) | (0.021) | (0.016) | (0.015) |
| experience | 0.040*** | 0.024** | 0.056*** | 0.045*** | 0.046*** |
| | (0.010) | (0.011) | (0.016) | (0.013) | (0.014) |
| education | 0.157*** | 0.144*** | 0.168*** | 0.133*** | 0.189*** |
| | (0.011) | (0.014) | (0.019) | (0.018) | (0.014) |
| english | 0.017 | -0.013 | 0.048** | 0.016 | 0.025 |
| | (0.010) | (0.013) | (0.018) | (0.015) | (0.014) |
| unemployed | -0.002 | -0.020 | 0.016 | 0.008 | -0.004 |
| | (0.009) | (0.012) | (0.014) | (0.014) | (0.012) |
| computer | 0.042*** | 0.055*** | 0.030* | 0.027* | 0.064*** |
| | (0.010) | (0.013) | (0.015) | (0.013) | (0.013) |
| wage | 0.003 | 0.006 | 0.000 | 0.017 | -0.003 |
| | (0.009) | (0.009) | (0.017) | (0.012) | (0.012) |
| samegender | 0.014 | 0.014 | 0.013 | -0.008 | 0.033 |
| | (0.018) | (0.024) | (0.019) | (0.024) | (0.023) |
| samegender × beauty | -0.040* | -0.042 | -0.035 | -0.025 | -0.054* |
| | (0.020) | (0.026) | (0.023) | (0.023) | (0.027) |
| time (photo) | 0.197 | 0.140 | 0.272 | 0.169 | 0.232 |
| | (0.126) | (0.168) | (0.164) | (0.177) | (0.153) |
| $2^{nd}$ position | 0.012 | 0.050 | -0.025 | -0.005 | 0.033 |
| | (0.017) | (0.030) | (0.029) | (0.026) | (0.030) |
| $3^{rd}$ position | 0.021 | 0.036 | 0.004 | -0.011 | 0.058** |
| | (0.022) | (0.033) | (0.028) | (0.032) | (0.024) |
| $4^{th}$ position | 0.039* | 0.087** | -0.009 | 0.000 | 0.075** |
| | (0.020) | (0.036) | (0.029) | (0.024) | (0.027) |
| Observations | 4384 | 2188 | 2196 | 2176 | 2208 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors clustered at the level of the applicant's photo are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

Table A4.6: Controls and beauty, interaction effects (twoway clustering by photo & participant)

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| Turkish origin | 0.031 | 0.004 | 0.063 | 0.103* | -0.047 |
| | (0.037) | (0.041) | (0.050) | (0.063) | (0.034) |
| headscarf | -0.197*** | -0.200* | -0.206*** | -0.123* | -0.239** |
| | (0.070) | (0.107) | (0.070) | (0.063) | (0.093) |
| headscarf × quality | 0.073** | 0.127*** | 0.022 | 0.039 | 0.118** |
| | (0.037) | (0.035) | (0.047) | (0.026) | (0.055) |
| headscarf × experience | 0.053*** | 0.046 | 0.060*** | 0.042** | 0.052*** |
| | (0.017) | (0.028) | (0.017) | (0.018) | (0.016) |
| headscarf × education | 0.039* | 0.082*** | 0.006 | -0.008 | 0.077** |
| | (0.022) | (0.024) | (0.027) | (0.032) | (0.037) |
| headscarf × english | 0.051*** | 0.023 | 0.076*** | 0.005 | 0.105*** |
| | (0.013) | (0.015) | (0.023) | (0.035) | (0.037) |
| beauty | 0.064*** | 0.063*** | 0.063*** | 0.050** | 0.077*** |
| | (0.015) | (0.019) | (0.023) | (0.025) | (0.019) |
| photo female | -0.046** | -0.079*** | -0.020 | -0.049 | -0.045* |
| | (0.020) | (0.019) | (0.035) | (0.040) | (0.025) |
| quality | 0.110*** | 0.100*** | 0.123*** | 0.175*** | 0.052*** |
| | (0.016) | (0.015) | (0.022) | (0.023) | (0.015) |
| experience | 0.043*** | 0.029** | 0.056*** | 0.051*** | 0.047*** |
| | (0.011) | (0.013) | (0.017) | (0.016) | (0.014) |
| education | 0.158*** | 0.140*** | 0.175*** | 0.142*** | 0.183*** |
| | (0.014) | (0.016) | (0.023) | (0.022) | (0.018) |
| english | 0.018 | -0.013 | 0.049*** | 0.017 | 0.024 |
| | (0.013) | (0.017) | (0.018) | (0.018) | (0.018) |
| unemployed | 0.003 | -0.020 | 0.026* | 0.013 | 0.002 |
| | (0.012) | (0.017) | (0.015) | (0.014) | (0.016) |
| computer | 0.045*** | 0.061*** | 0.032** | 0.027* | 0.073*** |
| | (0.011) | (0.014) | (0.016) | (0.016) | (0.013) |
| wage | 0.008 | 0.010 | 0.006 | 0.023* | 0.002 |
| | (0.011) | (0.010) | (0.019) | (0.014) | (0.014) |
| samegender | 0.021 | 0.027 | 0.013 | 0.003 | 0.035 |
| | (0.019) | (0.023) | (0.022) | (0.027) | (0.024) |
| samegender × beauty | -0.041** | -0.040 | -0.036 | -0.037 | -0.044* |
| | (0.020) | (0.026) | (0.027) | (0.023) | (0.025) |
| time (photo) | 0.224** | 0.129 | 0.352** | 0.241** | 0.208* |
| | (0.104) | (0.121) | (0.144) | (0.114) | (0.112) |
| $2^{nd}$ position | 0.002 | 0.039 | -0.034 | -0.006 | 0.012 |
| | (0.023) | (0.036) | (0.037) | (0.034) | (0.036) |
| $3^{rd}$ position | 0.023 | 0.047 | -0.003 | -0.002 | 0.051 |
| | (0.032) | (0.040) | (0.041) | (0.043) | (0.038) |
| $4^{th}$ position | 0.040* | 0.096*** | -0.017 | -0.002 | 0.082*** |
| | (0.022) | (0.036) | (0.032) | (0.028) | (0.030) |
| Observations | 3928 | 1960 | 1968 | 1944 | 1984 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardized beauty score of the photo. The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV. They are rescaled so that a higher value is always better for the employer. *samegender* is a dummy variables which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearence of a CV within the occupation. Robust standard errors two-way clustered at the level of the applicant's photo and participant are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

# Chapter 5

# Your wingman could help you land a job: How beauty composition of applicants affects the callback probability

# 5 Your wingman could help you land a job: How beauty composition of applicants affects the callback probability

*Co-authored with Attakrit Leckcivilize.*

*The economics department of Leibniz Universität Hannover provided a funding of 3,520 EUR. This funding body did not influence the experimental design. This funding is generally rewarded to cover implementation costs of studies in experimental economics at the Leibniz Universität Hannover.*

*Written in British English.*

## 5.1 Introduction

Does your appearance affect your earnings and career prospects? Hamermesh and Biddle (1994) were the first to use a nationally representative data set in the 1970s to estimate the effect of beauty on earnings in the US. They show that being rated as below average was associated with a wage penalty of 4% for women and 13% for men, while those with above average beauty received 8% and 4% in wage gain for women and men, respectively. Therefore, a shift in beauty rating from below average to above average is comparable to having an additional one and a half years of schooling.

Productive characteristics of appearance can only explain a fraction of the beauty premium in the labour market. Self-selection into occupations by appearance might arise for some occupations where workers have direct contact with customers or have to appear in public such as lawyers and salespersons. For example, using data of graduates from one law school, Biddle and Hamermesh (1998) showed that better-looking attorneys were more likely to work in private sector (where the interaction with clients, judges, and juries is higher). Hence, they earned more and the effect increased with experience. Yet some studies document a significant return to beauty in occupations where the productivity of appearance should be less pronounced such as American footballers and economics professors (see Hamermesh, 2011, for an extensive review).

One explanation for observing beauty premiums across all types of occupations is labour market discrimination against the less beautiful. On the one hand, the source of such prejudice might come from employer, employee, or customer discrimination. On the other hand, employers or recruiters have incomplete information about some unobservable characteristics of the applicant and hence they have to form some beliefs about such characteristics based on the “perceived” average of the group, which the applicant belongs to, such as ethnicity, or appearance (Altonji and Blank, 1999).

However, to estimate causal effects of appearance on labour market outcomes, researchers would ideally like to observe such outcomes of two groups which are the same in every aspect except for appearance. One technique to circumvent this issue is called “correspondence testing”. It explores the discrimination at the very first stage of getting into the labour market, i.e. the chance of being called back for a job interview (see, for example Bertrand and Mullainathan, 2004). The researcher has to create fake CVs and allocate the characteristic of interest such as appearance at random to each CV. On average, the CVs of these groups should be comparable in every observed characteristic except for their beauty.

Other than the beauty of oneself, does the appearance of other candidates influence the decision of recruiters or employers? Research studies in psychology, behavioural science, and marketing have shown that the preference ranking between two options in a choice set, which do not dominate another option in all dimensions, can be altered by adding irrelevant alternatives to the choice set. In the paradigm of asymmetrically dominated alternatives (Huber et al., 1982), the third alternative is constructed so that it is inferior (in every dimension) to one option but not the other one (Ariely and Wallsten, 1995). Hence, it is barely chosen as the best option. Yet the presence of this third alternative does influence the ranking between the other two relevant options, therefore, it is called a decoy.

Despite extensive research on the decoy effect in consumer research and marketing (see, for example Brenner et al., 1999; Slaughter et al., 2011; Lichters et al., 2015), this concept has been popularised and extended to dating and intimate relationships, for example, in TED talk by Dan Ariely[1]. He claimed that when one goes out bar-hopping, one will be more likely to be chosen for a date if he or she takes a friend, who looks similar to that person but slightly uglier, with him or her. This idea was dubbed as a wingman story. The wingman is the person who helps his or her friend to gain access to a romantic relationship or to avoid attention from whom they have no romantic interest. However, is this wingman concept with the decoy tweak applicable to the context outside of romantic relationships?

[1] Ariely, D. (2008, December). *Are we in control of our own decisions?* [Video file]. Retrieved from *https://www.ted.com/talks/dan_ariely_asks_are_we_in_control_of_our_own_decisions*

This paper combines a laboratory experiment with a randomised CV approach to assess the effects of own appearance and the beauty composition of candidates for the same job simultaneously. Even though existing literature used correspondence testing to analyse the role of body weight (Rooth, 2009) and the beauty premium (Kraft, 2012), field experiments with faked CVs do not allow researchers to control for beauty composition in the pool of applicants. We recruited students from local universities to participate in an experiment where they were asked to select applicants for an interview of fictitious positions from the pool of candidates whose CVs were randomised in every characteristic including their appearance. Further, we can control for personal characteristics of the “recruiters” and track the time our student participants used to evaluate the characteristics and photo page. We compute a relative time that the participant used on the photo page and use it as a proxy for “relative attention" to each photo in the job opening. Despite some applications in business and psychology, to the best of our knowledge, this experiment is the first to use such time-tracking technique together with the correspondence testing.

Our findings confirm that better-looking applicants get higher chances to be selected in all types of occupations, particularly in high-skilled jobs. This finding can reject neither productivity hypothesis of beauty nor taste based discrimination. Yet a larger effect of beauty in high skilled occupations supports a narrative of employee discrimination because in this sector our student participants could relate the applicants as their future co-workers. The effect of appearance is significantly higher if recruiters and candidates are of the opposite gender. With regards to beauty composition, we construct a wingman variable as an average beauty rating of other same-gender job applicants in the same job opening. Such a wingman effect is important in high-skilled jobs and jobs with customer contact. Arguably, candidates that are relatively more beautiful may help to increase sales in occupations with direct customer contact; however, the wingman effect in high-skilled sectors still supports the taste based discrimination narrative. This finding provides evidence for the decoy effect in the context of job application. Thus, it suggests that one can raise one’s chance to get a job by asking his or her less attractive friend(s), who are similar in other characteristics, to apply for the same job.

The paper is organised as follows: Section 5.2 explains the experimental design, Section 3 presents our methodology, Section 5.4 presents the results, Section 5.5 discusses the implications, and Section 5.6 concludes.

## 5.2 Experimental design

The main purpose of this paper is to fill the gap in the literature by showing that not only does personal appearance have a significant effect on invitation chances, but the beauty composition of other applicants of the same gender also matters. To answer this question, we conducted an experiment in order to randomly include different pictures of applicants to the similar CVs. Only in the setting of a lab experiment are we able to fully control not only for personal appearance, but also to construct our main independent variable, the appearance rating of a *wingman*. We define the *wingman* as someone of the same gender competing for the same job opening.

This paper uses data from the same experiment as described in Leckcivilize and Straub (2018). The experiment consists of two separate parts. First, we asked participants to act as HR staff and preselect candidates based on their CV for a job interview. In the second part, we asked participants of this experiment to rate the photos of the first part based on several characteristics including appearance. The beauty variable in our analysis was constructed with data from the second part of the experiment.

The "recruiting process" in the first part was conducted in December 2015; 120 students participated in 9 sessions. The sessions lasted one hour and reimbursement was 20 Euro. Our participants reviewed stylised CVs, each of these included a photo[2], a randomly assigned first and last name, and seven characteristics of a CV which are reported in Table 5.1. We framed the task such that all applicants fulfil the formal conditions for the respective job and differ only by their CVs. Each participant reviewed 32-48 CVs for 8-12 jobs, organised in 2-3 blocks[3] of four positions. In March 2016, we conducted the second part of the experiment with 40 students with the objective to acquire a beauty rating of the photos which we used in the first part.

We used a three step procedure in order to randomly assign photos, names, and characteristics to each CV. In the first step we restricted the possible combinations of characteristics to 834, so that the chance that one CV would dominate the others in every characteristic is almost zero. After assigning one of these possible combinations we assigned photos to these CV randomly. Thereby we ensured that no photo would appear twice in each block. In the third step we randomly picked

[2]It is a common practice in Germany to put an applicant's photo on a CV.

[3]Separated by a paper-and-pencil questionnaire about socio-demographic characteristics and a questionnaire for another project.

Table 5.1: Characteristics in a CV (in the presented order)

| | Categories | Min | Max |
|---|---|---|---|
| working experience | 4 | 0 | 3 |
| asked wage | 5 | 20% lower than average | 20% higher than average |
| final grade | 3 | lower than average | higher than average |
| reputation of the college (only high skilled jobs) | 3 | lower than average | higher than average |
| amount of absence days (only low skilled jobs) | 3 | lower than average | higher than average |
| current unemployment | 4 | currently not employed | 13-18 month |
| computer skills | 3 | sufficient | very good |
| English skills | 3 | basic | fluent |

*Notes:* We refer to reputation of the college and amount of absence days as *quality* in our regressions, the latter one is reported on a school certificate and is a decisive hiring criteria for low skilled job entrants.

a name for each photo. One of 150 first names and 100 last names was randomly assigned, so that it fitted the gender and ethnicity of the photo.

Our experimental design allows to group professions into low skilled and high skilled occupations as well as jobs with customer contact and without in order to analyse productive characteristics of appearance (Biddle and Hamermesh, 1998). Concerning the beauty rating in the second part we used "double standardisation" (within photos & raters) before we assign those to the CVs. Additionally, we examine the time spent looking at the photo by separating the photos and characteristics (see Figure A5.1 and Figure A5.2) in order to examine whether this is correlated with discrimination (Bartoš et al., 2016).

## 5.3 Methodology

### 5.3.1 Theoretical background

In this section we show how appearance in a taste discrimination framework developed by Becker (1957) influences labour market outcomes in a neoclassical setting (see, for example Hamermesh, 1986). More specifically, we describe that own appearance should affect chances to be selected for an interview through two channels because it increases productivity and decreases taste based discrimination. Additionally, we hypothesise how the beauty composition of other applicants of the same gender affects the probability to be chosen.

Assuming perfect labour market competition with constant wages $w$ and imperfect competition on the product market, where firms can influence the price of the output. Firm's profit $\Pi(L)$ is a function of quantity $q(L)$ times the price of the product $p(q(L))$, labour costs $w \times L$, where $L$ is the amount of workers, and other costs $c$:

$$\Pi(L) = p(q(L)) \times q(L) - w \times L - c \tag{5.1}$$

If $c$ is constant in the short run a profit maximising firm would only hire an additional worker $i$ if $\Pi'(L) > 0$, i.e:

$$MRP_L \geq w \tag{5.2}$$

where $MRP_L$ is the first derivative of $p(q(L)) \times q(L)$ with respect to $L$ and represents the marginal revenue product of labour, i.e. the incremental change in revenues by employing one additional worker.[4] We further assume that there exist perceived wages, which can be expressed by the actual wage and a discrimination factor which differs by worker and firm, and is among others a function of appearance. Also we suppose that productivity is not constant over the workers, but is also a function of beauty and other characteristics of worker $i$. We define $\tilde{w}^i$ as the perceived wage of worker $i$ as a function of the actual wage $w^i$ and a discrimination factor $d_j$ of firm $j$ which is a function of person $i$ beauty rating, the beauty composition of other applicants of the same gender, and other discriminatory factors $z$ and can be specified as:

$$\tilde{w}^i \equiv w^i(1 + d_j(z_i, V[beauty_i, wingman_{ik}])) \tag{5.3}$$

where $V[beauty_i, wingman_{ik}]$ is a composite function for the effects of own ($i$) and wingman's beauty in the job opening $k$ on the discrimination factor $d_j$, with $\frac{\partial d_j}{\partial beauty_i} < 0$. This term $d_j$ captures the non-pecuniary mark-up of hiring a person with, e.g. a certain beauty rating. Suppose that beauty has both a productive quality and a discrimination aspect, and therefore increases the $MRP_L$ and decreases the perceived wage $\tilde{w}$. Allowing for worker heterogeneity then transforms equation (5.2) to:

$$MRP_L^i \geq \tilde{w}^i \tag{5.4}$$

[4] $MRP_L$ can be further decomposed into the marginal revenue ($MR = \frac{\partial p(q) \cdot q}{\partial q}$) times the marginal product of worker $i$ ($MP_L = \frac{\partial q}{\partial L}$), hence, $MRP_L = MR \times MP_L = \frac{\partial p(q) \cdot q}{\partial q} \cdot \frac{\partial q}{\partial L} = \frac{\partial p(q(L)) \cdot q(L)}{\partial L}$.

Additionally, the marginal revenue product of labour is also a function of other productive characteristics $x_i$. In this case the the hiring condition for any worker $i$ can be expressed as:

$$Prob(MRP_L^i \geq \tilde{w}^i) = Prob(MRP_L^i(x_i, beauty_i) \geq w^i(1 + d_j(z_i, V[beauty_i, wingman_{ik}]))) \quad (5.5)$$

In other words, appearance of the candidates can affect the probability to be chosen through two complementing channels, i.e. potential productivity enhancing through, for example, higher sales and through a reduction of taste based discrimination in favour of more attractive candidates.[5] However, as beauty may not be perfectly measured and compared, we hypothesise that the beauty composition of the pool of candidates could potentially influence the discrimination factor by aiding a comparison between candidates as a wingman or decoy option.[6] Therefore, we expect that the 'informative' value of the wingman should be reduced as the recruiters are more experienced and a decoy option should not directly affect productivity.

### 5.3.2 Empirical strategy

We adopt the Linear Probability Model (LPM) to estimate the impacts of own beauty and the beauty composition of the whole pool of applicants on the chance of being selected for an interview by the participants in our experiment.

$$y_{ijk} = \alpha + B'_{jk}\beta + X'_i\gamma + Z'_{ij}\theta + \varepsilon_{ijk}$$

where $y_{ijk}$ is a dummy variable equal to one if CV $i$ with photo $k$ is chosen by participant $j$. $B'_{jk}$ is a vector for our main explanatory variables, which are a composite beauty rating score of photo $k$ and the beauty composition of the pool of applicants as well as control variables for gender, having the same gender as recruiter $j$, ethnicity, and wearing a headscarf. $X'_i$ is a vector of the CV's seven characteristics discussed in Section 5.2 and control variables for the order of appearance of the CV.[7] $Z'_{ij}$ is a vector of participant $j$ characteristics from the participant's

[5] In our experiment, we asked the participants to choose a fixed number of candidates (two out of four) for an interview. Hence, the participants could not choose any numbers of candidates satisfying this condition. Therefore, our participants just chose two candidates with the highest value of $MRP_L^i - \tilde{w}^i$.

[6] If the wingman also provides extra information on the distribution of beauty among candidates such as in a job opening with few applicants, it is likely that the $wingman_{ik}$ could affect the marginal revenue product of labour ($MRP_L^i$) indirectly through a 'better' decision of the recruiter based on a 'better' beauty ranking.

[7] These dummy variables for the order of CV $i$, i.e. applicant number 2, 3 or 4 in each job-position (with the first applicant as the reference group) are included to control for a tendency that some participants might systematically choose the first, second, third or fourth applicant more often than others.

responses in a questionnaire and the relative time spent on the photo of CV $i$ with respect to the time spent on all candidates for that job opening.

Within $B'_{jk}$ we assess an effect of beauty rating when the applicant is of the same gender as the recruiter (student participant) by adding an interaction term between the beauty score and a dummy *samegender* equal to one if the recruiter and the applicant have the same gender. Moreover, the beauty composition of applicants for the same job opening (*wingman beauty*) is measured by an average beauty score of all other applicants in that job opening with the same gender as the applicant $k$. For example, if there are three female and one male applicant for the same job, the *wingman beauty* of each female applicant is equal to an average standardised beauty of the other two female applicants, while for the only male applicant to this job, his *wingman beauty* would be zero.

Since our variables of interest are drawn from the photo accompanying each CV, we also control for proportion of time each participant $j$ looked at the photo page of CV $i$ ($time_{ij}$, which is a component in $Z'_{ij}$). Controlling for the relative time each participant used to inspect the photo might help to address a concern on attention discrimination proposed by Bartoš et al. (2016). Their study found that minority names reduce recruiters' effort to check the resume. Instead of being a minority, having less attractive photos might affect participants' effort to look and review the CV. Therefore, our results can be interpreted as beauty effects given the same level of effort by the recruiter.

We cluster the error terms $\varepsilon_{ijk}$ on both photo and participant level (Two-way clustered robust standard errors, (Cameron and Miller, 2015)). As robustness checks, we also perform one-way clustered standard errors either by participant or by photo. Lastly, in order to ascertain whether the results are driven by any particular characteristics of job openings or participants, we estimate the model for several sub-samples based on job classifications, comprising four types of jobs (high skilled, low skilled, with and without customer contact), and the gender of participants.

## 5.4 Results

Our results from linear probability models with two-way clustered standard errors by photo and participant, as explained in Cameron and Miller (2015) are presented. We later perform robustness checks with different clustering options. Starting with a model that includes only beauty and a set of dummy variables for the position of the CV in the job opening, the first column of Table 5.2 shows a highly significant effect of personal appearance on the probability to be chosen. When we add linear control variables for characteristics of the corresponding CV in the second column of Table 5.2, the significance level of the beauty premium remains unchanged. Table A5.2 shows the results of the second column of Table 5.2 by type of occupation and the estimated coefficients for the CV's characteristics, revealing that the coefficients of the candidates characteristics have the expected sign. The results in Table A5.2 are also in line with existing literature and theoretical prediction that the beauty premium should be more pronounced in contact jobs. Additionally, we observe a positive and significant beauty premium in high skilled jobs.

Conducting correspondence testing in the laboratory enables us to include characteristics of not only the corresponding CV, but also from the participants who made the decision. In order to control for heterogeneity in perceived beauty between a participant with the same gender to the CV's photo and those of opposite gender, we construct a dummy variable *samegender*, which is equal to one if the decision maker and the applicant are of the same gender and zero otherwise. The third column of Table 5.2 presents the results from regressions where this variable as well as an interaction term between beauty and *samegender* are included. Such an inclusion increases the magnitude of the beauty effect on the chance of being selected. It indicates that the observed beauty premium in the first and second column are mainly driven by the effects from the opposite gender, while the beauty effects from the same gender are quite close to zero in all occupation types. Furthermore, including this interaction term exhibits a significant beauty premium (at the 5% level) for the opposite gender in no-contact jobs.

Besides German looking photographs, we included Turkish looking ones in order to mimic the real demographic composition of job candidates in Germany. Additionally, we acquired photos of the same Turkish women with headscarves and without to test potential discrimination on religious practice. In the third column of Table 5.2 we present results after including these dummy variables for photo characteristics. Ethnicity and headscarf do not affect the average

Table 5.2: The effect of beauty & wingman beauty (all occupations).

| Type of Occupation | All | All | All | All | All |
|---|---|---|---|---|---|
| beauty | 0.025*** | 0.024*** | 0.061*** | 0.060*** | 0.060*** |
| | (0.009) | (0.008) | (0.015) | (0.014) | (0.015) |
| $2^{nd}$ position | 0.016 | 0.015 | 0.012 | 0.014 | 0.012 |
| | (0.025) | (0.022) | (0.022) | (0.022) | (0.023) |
| $3^{rd}$ position | 0.035 | 0.025 | 0.020 | 0.020 | 0.018 |
| | (0.028) | (0.028) | (0.027) | (0.027) | (0.028) |
| $4^{th}$ position | 0.047** | 0.041* | 0.038* | 0.039* | 0.039* |
| | (0.022) | (0.021) | (0.021) | (0.021) | (0.022) |
| beauty × samegender | | | -0.039* | -0.038* | -0.039* |
| | | | (0.020) | (0.020) | (0.020) |
| samegender | | | 0.015 | 0.015 | 0.015 |
| | | | (0.018) | (0.018) | (0.018) |
| photo female | | | -0.050*** | -0.035* | -0.031 |
| | | | (0.017) | (0.019) | (0.019) |
| Turkish origin | | | 0.045 | 0.049 | 0.048 |
| | | | (0.036) | (0.035) | (0.035) |
| headscarf | | | 0.013 | 0.016 | 0.016 |
| | | | (0.039) | (0.041) | (0.041) |
| wingman beauty | | | | -0.043* | -0.041* |
| | | | | (0.023) | (0.023) |
| time (photo) | | | | | 0.098 |
| | | | | | (0.314) |
| $(\text{time (photo)})^2$ | | | | | 0.300 |
| | | | | | (0.987) |
| # samegender applicants | | | | | -0.012 |
| | | | | | (0.008) |
| Constant | 0.475*** | 0.097 | 0.451*** | 0.443*** | 0.459*** |
| | (0.015) | (0.071) | (0.039) | (0.040) | (0.038) |
| Characteristics | No | Linear | Categorical | Categorical | Categorical |
| Observations | 4,384 | 4,384 | 4,384 | 4,384 | 4,384 |
| Adj. $R^2$ | 0.004 | 0.078 | 0.090 | 0.091 | 0.092 |

*Notes:* This table shows the relationship between applicant's appearance and position of the CV and the chances that the CV is selected for a job interview. *beauty* is a double-standardised beauty score of the photo. *wingman beauty* is the average *beauty*-score of the applicants of the same gender for the same position. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *Turkish origin*, *headscarf*, and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *position* refers to the order of appearance of a CV within the occupation (with the first applicant as the reference group). *time(photo)* is the relative time each participant used to inspect the photo. *# male applicants* is the number of male applicants in the job, ranging from 0 to 4 (see page 90 for a detailed discussion of the variables). Robust two-way clustered (by participant & photo) standard errors in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

probability to get chosen, but the headscarf does matter when interacted with CV characteristics as discussed in Leckcivilize and Straub (2018). We also include a dummy variable which is equal to one if the applicant is female. The coefficient is negative and significant. This might be driven by the fact that female photos in our sample were rated as more beautiful than male photos on average. We also replace linear controls for the characteristics with categorical controls so as to allow for non-linearity in the effect of characteristics of our CVs. These models have higher adjusted $R^2$ but are similar in the magnitude and significance level of the coefficients of interest. Therefore we use categorical controls in all following Tables.

In the fourth column of Table 5.2 we add our main variable of interest, which is the wingman beauty. Its coefficient is negative and significant at the 10% level for the whole sample and at the 5% level for contact and high-skilled jobs. For both high-skilled and contact occupations, the size of the effects are comparable and the probabilities to get chosen reduce by roughly 6.5 percentage points if the average beauty of the same gender competitors increases by one standard deviation (see table A5.3). The magnitude is quite remarkable given that each participant has to choose two out of four applications in each job opening, hence, the ex-ante chance to be chosen for any candidates is 50%. More importantly, the coefficients for beauty in the fourth column of Table 5.2 remain almost unchanged compared to the third column. This implies that the wingman effect does not crowd out the beauty premium but adds on top of it.

We add further controls for the number of same gender applicants and the relative time a decision maker stays on the photo page, compared to the time she spends on reviewing all applicants in that job position. This relative time variable does not influence the result as presented in fifth column of Table 5.2 and in Table A5.3 by type of occupation. When clustering standard errors only by participant and not by participant and photo, the standard errors get even smaller. As shown in Table A5.4, it results in significant effects of the wingman variable at the 1% level for both contact and high-skilled occupations.

Table 5.3: Separate regressions for $1^{st}$ and $2^{nd}$ choice

| Type of Occupation | All | | No Contact | | Contact | | Low Skill | | High Skill | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice |
| beauty | 0.040*** | 0.045*** | 0.021* | 0.060*** | 0.057*** | 0.032 | 0.026 | 0.031 | 0.047** | 0.057** |
| | (0.011) | (0.015) | (0.012) | (0.017) | (0.015) | (0.024) | (0.020) | (0.031) | (0.019) | (0.023) |
| wingman beauty | -0.022 | -0.039* | -0.007 | -0.014 | -0.036 | -0.070* | -0.010 | -0.022 | -0.035* | -0.054** |
| | (0.016) | (0.023) | (0.018) | (0.030) | (0.024) | (0.037) | (0.018) | (0.027) | (0.021) | (0.028) |
| beauty × samegender | -0.026* | -0.027 | -0.010 | -0.039 | -0.038* | -0.015 | -0.029 | -0.010 | -0.024 | -0.041 |
| | (0.014) | (0.020) | (0.018) | (0.027) | (0.021) | (0.024) | (0.018) | (0.023) | (0.019) | (0.028) |
| samegender | 0.017 | 0.005 | 0.013 | 0.005 | 0.022 | 0.001 | 0.015 | -0.023 | 0.019 | 0.026 |
| | (0.016) | (0.019) | (0.017) | (0.029) | (0.020) | (0.023) | (0.018) | (0.025) | (0.024) | (0.027) |
| photo female | -0.014 | -0.033 | -0.023 | -0.082*** | -0.006 | 0.015 | 0.015 | -0.068 | -0.036 | 0.004 |
| | (0.016) | (0.023) | (0.019) | (0.022) | (0.026) | (0.036) | (0.040) | (0.053) | (0.025) | (0.034) |
| Turkish origin | 0.045** | 0.032 | 0.002 | 0.038 | 0.093** | 0.033 | 0.064 | 0.080 | 0.012 | -0.017 |
| | (0.021) | (0.039) | (0.021) | (0.045) | (0.038) | (0.056) | (0.048) | (0.076) | (0.028) | (0.045) |
| headscarf | -0.017 | 0.026 | -0.001 | 0.089 | -0.041 | -0.044 | -0.044 | -0.014 | 0.020 | 0.069 |
| | (0.030) | (0.044) | (0.039) | (0.056) | (0.030) | (0.043) | (0.043) | (0.048) | (0.022) | (0.047) |
| $2^{nd}$ position | 0.009 | 0.014 | 0.050 | 0.024 | -0.032 | 0.002 | -0.026 | 0.017 | 0.042 | 0.008 |
| | (0.025) | (0.023) | (0.033) | (0.036) | (0.034) | (0.033) | (0.031) | (0.035) | (0.035) | (0.029) |
| $3^{rd}$ position | -0.002 | 0.031 | 0.026 | 0.024 | -0.034 | 0.034 | -0.017 | -0.003 | 0.008 | 0.065* |
| | (0.022) | (0.030) | (0.030) | (0.042) | (0.030) | (0.036) | (0.033) | (0.037) | (0.031) | (0.039) |
| $4^{th}$ position | -0.090*** | 0.125*** | -0.065** | 0.159*** | -0.115*** | 0.081** | -0.116*** | 0.093*** | -0.071** | 0.149*** |
| | (0.024) | (0.023) | (0.030) | (0.038) | (0.027) | (0.035) | (0.031) | (0.024) | (0.031) | (0.034) |
| 1 year of experience | 0.085*** | 0.040 | 0.078** | 0.045 | 0.089*** | 0.033 | 0.081*** | 0.035 | 0.095*** | 0.055 |
| | (0.015) | (0.026) | (0.032) | (0.031) | (0.016) | (0.034) | (0.028) | (0.036) | (0.026) | (0.037) |
| 2 years of experience | 0.108*** | 0.076** | 0.092*** | 0.033 | 0.125*** | 0.115** | 0.111*** | 0.083* | 0.111*** | 0.081** |
| | (0.020) | (0.029) | (0.031) | (0.030) | (0.026) | (0.055) | (0.030) | (0.044) | (0.024) | (0.037) |
| 3 years of experience | 0.130*** | 0.043 | 0.133*** | 0.008 | 0.126*** | 0.078 | 0.138*** | 0.067 | 0.144*** | 0.051 |
| | (0.025) | (0.038) | (0.036) | (0.046) | (0.034) | (0.059) | (0.033) | (0.047) | (0.036) | (0.048) |
| asked wage: 20% more than average | -0.040 | -0.024 | -0.066* | -0.037 | -0.021 | -0.011 | -0.094** | -0.062* | -0.002 | 0.004 |
| | (0.029) | (0.029) | (0.039) | (0.043) | (0.026) | (0.059) | (0.040) | (0.037) | (0.031) | (0.045) |
| asked wage: 10% more than average | -0.011 | -0.032 | -0.069*** | -0.057 | 0.046 | -0.001 | -0.071*** | -0.049 | 0.036 | -0.014 |
| | (0.022) | (0.026) | (0.027) | (0.039) | (0.037) | (0.045) | (0.022) | (0.033) | (0.036) | (0.041) |
| asked wage: 10% less than average | -0.036* | -0.016 | -0.058** | -0.015 | -0.013 | -0.014 | -0.057* | -0.010 | -0.013 | -0.020 |
| | (0.019) | (0.022) | (0.027) | (0.024) | (0.029) | (0.038) | (0.032) | (0.031) | (0.024) | (0.026) |
| asked wage: 20% less than average | -0.048*** | -0.021 | -0.056* | -0.044 | -0.035 | 0.005 | -0.061** | -0.029 | -0.026 | -0.007 |
| | (0.018) | (0.033) | (0.031) | (0.040) | (0.023) | (0.055) | (0.031) | (0.042) | (0.027) | (0.041) |
| education: below average | -0.122*** | -0.122*** | -0.105*** | -0.164*** | -0.141*** | -0.082** | -0.112*** | -0.130*** | -0.141*** | -0.115*** |
| | (0.015) | (0.026) | (0.017) | (0.027) | (0.020) | (0.036) | (0.020) | (0.036) | (0.020) | (0.035) |
| education: above average | 0.091*** | 0.110*** | 0.091*** | 0.068** | 0.092*** | 0.158*** | 0.065** | 0.068*** | 0.123*** | 0.159*** |
| | (0.020) | (0.018) | (0.024) | (0.030) | (0.025) | (0.026) | (0.028) | (0.023) | (0.023) | (0.031) |
| quality: below average | -0.121*** | -0.130*** | -0.122*** | -0.123*** | -0.119*** | -0.142*** | -0.182*** | -0.182*** | -0.060*** | -0.076*** |
| | (0.011) | (0.020) | (0.020) | (0.023) | (0.020) | (0.032) | (0.020) | (0.033) | (0.017) | (0.024) |
| quality: above average | 0.051*** | 0.011 | 0.077*** | -0.010 | 0.027 | 0.030 | 0.080*** | 0.035 | 0.032 | 0.005 |
| | (0.017) | (0.023) | (0.021) | (0.031) | (0.019) | (0.033) | (0.023) | (0.036) | (0.022) | (0.021) |
| 13-18 month unemployed | 0.014 | 0.016 | 0.036 | 0.064 | -0.011 | -0.030 | 0.019 | -0.038 | -0.011 | 0.062 |
| | (0.033) | (0.028) | (0.048) | (0.042) | (0.040) | (0.055) | (0.041) | (0.044) | (0.042) | (0.051) |
| 7-12 month unemployed | -0.015 | 0.026 | 0.000 | 0.058 | -0.033 | -0.012 | 0.001 | -0.038 | -0.044 | 0.078* |
| | (0.026) | (0.022) | (0.041) | (0.044) | (0.025) | (0.043) | (0.036) | (0.033) | (0.031) | (0.041) |
| 1-6 month unemployed | -0.019 | 0.004 | 0.008 | 0.031 | -0.044 | -0.024 | -0.033 | -0.012 | -0.014 | 0.017 |
| | (0.022) | (0.017) | (0.033) | (0.032) | (0.028) | (0.029) | (0.031) | (0.031) | (0.026) | (0.030) |
| computer skills: good | 0.029 | 0.037 | 0.050* | 0.071** | 0.009 | 0.005 | 0.033 | 0.013 | 0.032 | 0.066* |
| | (0.020) | (0.024) | (0.029) | (0.032) | (0.023) | (0.028) | (0.026) | (0.028) | (0.022) | (0.037) |
| computer skills: very good | 0.050*** | 0.044** | 0.067*** | 0.068** | 0.043* | 0.024 | 0.025 | 0.037 | 0.089*** | 0.068*** |
| | (0.016) | (0.022) | (0.021) | (0.028) | (0.025) | (0.031) | (0.020) | (0.032) | (0.023) | (0.023) |
| English: advanced | 0.045** | 0.017 | 0.007 | -0.020 | 0.089*** | 0.055** | 0.020 | 0.003 | 0.065*** | 0.029 |
| | (0.017) | (0.020) | (0.022) | (0.031) | (0.020) | (0.026) | (0.027) | (0.027) | (0.021) | (0.024) |
| English: fluently | 0.005 | 0.026 | -0.033 | -0.019 | 0.045** | 0.068* | 0.008 | 0.020 | 0.011 | 0.044* |
| | (0.015) | (0.022) | (0.025) | (0.030) | (0.018) | (0.039) | (0.021) | (0.034) | (0.022) | (0.026) |
| Observations | 4384 | 3288 | 2188 | 1641 | 2196 | 1647 | 2176 | 1632 | 2208 | 1656 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview as $1^{st}$ *Choice* or $2^{nd}$ *Choice* respectively. The number of observations for $2^{nd}$ *Choice* decreases because CV's which were selected as first preference are excluded from these estimations. *Turkish origin*, *headscarf*, and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardised beauty score of the photo. The variables *quality (base: average)*, *experience (base: 0 years of experience)*, *education (base: average)*, *english (base: basic)*, *unemployed (base: currently not unemployed)*, *asked wage (base: average wage)*, and *computer (base: sufficient)* refer to the characteristics of each CV. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearance of a CV within the occupation. Robust standard errors two-way clustered (by photo & participant) are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

We also utilise the ranking of first and second choice in our experiment and present the results in Table 5.3.[8] The beauty effects are robustly positive and significant for both first and second choice in all jobs and high-skilled jobs. Interestingly, appearance is one of the important determinants of the first choice in contact jobs. This might signal the potential productivity enhanced characteristic of beauty in occupations with more customer contact. Furthermore, appearance is only positive and significant at the 1% level in the second choice of jobs with less customer contact, hinting some discriminating effects at the border in favour of more beautiful applicants. Regarding the wingman effect, it appears predominantly among the second choice. Therefore, the extra impact of relative beauty to the same gender competitors seems to kick in when the decision is at the borderline, i.e. those who barely made it rather than the favourite candidate.

Separated regressions by gender of the decision maker with the first and second choice are shown in Table A5.5 and A5.6. They show qualitatively similar effects of other characteristics such as work experience, education, quality, and English skills between genders. Significant effects of the wingman in all sample, contact, and high-skilled jobs are driven mainly by male participants, whereas female participants reward more on absolute level of beauty rating. Moreover with this separation by gender, our sample size is not large enough to precisely distinguish the beauty effects of many job categories. Similar to Table 5.3, both the beauty premium and wingman effect prevail predominantly in the second choice. Thus, the beauty composition significantly contributes to the decision making at the margin.

Furthermore, our participants were asked to answer the Cognitive Reflection Test (CRT) proposed by Frederick (2005).[9] We use the CRT as a measure for the participant's ability to pause and reflect on a question before carefully responding to it. In our experiment, we hypothesise that the participants who answered all CRT questions correctly are more likely to make their decision based on system 2, or the deliberate system rather than system 1, or the intuitive system (Kahneman and Egan, 2011). However, since our experiment was not incentivised (and probably not straightforward to do so), we interpret that participants with CRT score of less than 4 are a mixture of those who chiefly rely on their system 1 and those who put less effort into our experiment.

[8] In the first choice regression, we assign the dummy $y_{ijk}$ to be one only if CV $i$ with photo $k$ is chosen by participant $j$ as the first choice and zero otherwise. As for the second choice regression, we drop the first choice, so the dummy $y_{ijk}$ is equal to one only for the selected second choice.

[9] Our test consists of 3 originally questions presented in Frederick (2005) and 1 adapted question.

Table A5.7 shows the results by those participants who answered all 4 CRT questions correctly and those who did not. Our main findings that the beauty premium and wingman effect prevail in high skilled occupations and jobs with extensive customer contact are reiterated only among the participants who answered all CRT questions correctly but not those who got some question(s) wrong. Though it might be interpreted as the participants who rely on the intuitive system are less biased towards the beautiful, this is a bit far-fetched because the CRT is an imperfect measure capturing intuition-led respondents as well as random clickers. Our focus is rather on the group who answered all CRT questions correctly. If this is a proxy for those who rely more on the deliberate system to make their decision, our finding shows that both beauty premium and wingman effect are not solely a result of an over-reliance on intuition. Rather, the wingman (or decoy) effect still exists even among decision makers with deliberative thought processes.

Finally, Table A5.8 presents the results based on sub-sample of observations from the first block versus the second block of the experiment. Beauty premiums are statistically significant at the 10% level in all job types from the first block, while the only appearance of candidates in high skilled occupations still matter in the second block (the latter part of our experiment). However, wingman beauty is important for the selection only in high skilled and contact jobs from the first block.

## 5.5 Discussion

Why does absolute and relative level of beauty as proxied by the beauty rating and the wingman variable matter for the chance to be chosen? And why do such effects differ by recruiters' characteristics and the ranking among selected candidates? First, attractive workers in customer oriented occupations could raise the productivity of the firm by, for example, increasing its sales. Moreover, better-looking people may be substantially more self-confident than others, hence improving their performance (see Mobius and Rosenblat, 2006; Cipriani and Zago, 2011). Our result in Table 5.3 supports the productivity hypothesis as the beauty premium is significant for the first choice in the occupations with more customer contact.

Yet a beauty premium could result from taste based discrimination as well (Becker, 1957). For instance, the significant effect in high skilled occupations could be circumstantial evidence of employee discrimination because all participants are college students and they can probably identify themselves as potential future co-workers of the candidates in high skilled professions.

Interestingly, such a discrimination in favour of the good-looking becomes important at the margin for occupations with less customer contact where beauty has no clear implication on productivity.

Regarding differential effects of beauty composition of other candidates by recruiters' genders, our finding that the wingman effect is driven by male participants is in line with impacts of peer appearance and students' grade by (Hernández-Julián and Peters, 2018). They found that students, especially female, in courses filled with attractive classmates received significantly higher grades from younger and male instructors than those students with average-looking peers. In their context, exam grades of the whole class are not a zero-sum game, therefore, better peer appearance can benefit everyone in the class. However, in the job application setting, competing with a more attractive wingman makes the candidate worse off.

Are beauty and wingman effects mainly prominent among those relying more on intuitive (system 1) decision making? Separated regressions by Cognitive Reflection Test suggest that decision makers with either high or low CRT do take basic candidates' characteristics into their consideration but only the high CRT group gives extra weight on own and the wingman's beauty. Therefore, such a finding does not support an argument that system 1 is a sole underlying mechanism of the beauty and wingman effects. Rather, the role of appearance seems to be more complex and deep-rooted even among those with a more deliberative decision making process.

Is it possible that the wingman's beauty helps to provide extra information to the decision makers? To get some evidence for this hypothesis, we analyse whether the effects of wingman differ by the participants' level of familiarity with the tasks by running separated regressions between the first and second block of the main experiment. The wingman effect is significant at the 5% level in high skilled and customer oriented occupations from the first block only. In other words, after the participants went through four job openings in the first block of the experiment and became familiar with the tasks and our photos (we randomly inserted 29 photos into 32 - 40 CVs, hence a couple of photos can appear twice), the wingman effect and to a lesser extent, the beauty premium seem to fade out.

This is in line with our hypothesis that the wingman works like a decoy option, which provides 'extra' information for a comparison between different candidates. In our case, since beauty level might not be precisely measured, the wingman could help the recruiters to gather information about the beauty distribution of all candidates (either as a signal for productivity, or for discrimination purposes). Moreover, as the wingman effects are driven by the recruiters who used the deliberative system 2 for their selection process, we can infer that this group of recruiters is the one that requires more information on the beauty distribution, hence uses the wingman beauty in the first block. Yet once they form some beliefs on the beauty distribution of the candidates, in the second block wingman beauty does not provide significantly more information on beauty ranking anymore. Thus, the wingman variables are of lower magnitude and become insignificant. Nonetheless, the potential discrimination aspect of beauty is still persistent in high skilled jobs.

## 5.6 Conclusion

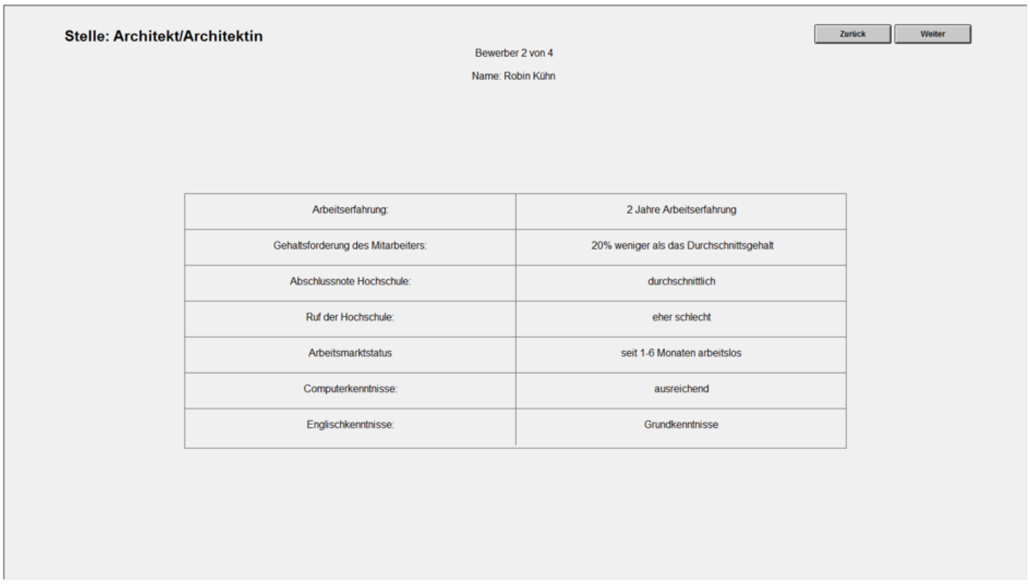
Our results indicate that the mechanism of the decoy effect also influences the hiring process. We observe a strongly significant decoy effect only among male decision makers. Although both genders value appearance, only males seem to compare it to the relative beauty ratings of other contestants. The magnitude of the beauty premium is relatively high and it is robust to an inclusion of the wingman's beauty score. For example, suppose that two women apply for a high skilled occupation. One of them has one standard deviation higher beauty rating than the average, while the other's appearance is one standard deviation lower. Let their CVs have average quality and both of them are evaluated by a male recruiter. Based on our experiment, the chance for the unattractive candidate to be chosen is only 32%, whereas the attractive applicant has a 68% chance to be chosen, or more than double the odds.

We are able to show that the rich body of literature which documents the impact of appearance in the labour market might underestimate the true impact of beauty due to the additional impact of the wingman beauty. Similar to a decoy option, the wingman provides more 'relevant' information on the beauty distribution. Further, according to our results between the groups with low and high CRT scores, the information hypothesis seem to be more plausible than the narrative about 'superficial' recruiters who are victims of their unconscious biases. Hence, the decoy effect could influence hiring decisions, especially in professions where the pool of candidates is small or the gender ratio is very unequal. Conversely, the larger the pool of candidates with the same gender (or other demographic group) is, the smaller variation the wingman can induce.

## 5.7 Appendix

Figure A5.1: Example of the experiment screen 1


Figure A5.2: Example of the experiment screen 2




| | |
|---|---|
| Arbeitserfahrung: | 2 Jahre Arbeitserfahrung |
| Gehaltsforderung des Mitarbeiters: | 20% weniger als das Durchschnittsgehalt |
| Abschlussnote Hochschule: | durchschnittlich |
| Ruf der Hochschule: | eher schlecht |
| Arbeitsmarktstatus | seit 1-6 Monaten arbeitslos |
| Computerkenntnisse: | ausreichend |
| Englischkenntnisse: | Grundkenntnisse |

Table A5.1: Descriptive Statistics for the experiments

| | Mean | SD | Min | Percentile | | | Max |
|---|---|---|---|---|---|---|---|
| | | | | 25th | 50th | 75th | |
| **Panel A: Main experiment** | | | | | | | |
| female | 0.463 | 0.501 | | | | | |
| age | 23.280 | 3.234 | 18 | 21 | 23 | 24 | 45 |
| siblings | 1.361 | 0.989 | 0 | 1 | 1 | 2 | 4 |
| study semester | 3.983 | 2.626 | 1 | 2 | 3 | 5 | 15 |
| correct CRT answers | 3.100 | 1.088 | 0 | 2.5 | 3 | 4 | 4 |
| bachelors | 0.600 | 0.492 | | | | | |
| masters | 0.283 | 0.453 | | | | | |
| born in Germany | 0.941 | 0.236 | | | | | |
| migration background | 0.192 | 0.395 | | | | | |
| **Panel B: Beauty rating** | | | | | | | |
| female | 0.525 | 0.506 | | | | | |
| age | 23.700 | 3.291 | 18 | 21 | 23 | 25 | 33 |
| siblings | 1.600 | 0.955 | 0 | 1 | 2 | 2 | 4 |
| study semester | 4.900 | 3.233 | 1 | 2 | 5 | 7 | 15 |
| correct CRT answers | 2.925 | 1.118 | 0 | 2 | 3 | 4 | 4 |
| bachelors | 0.575 | 0.501 | | | | | |
| masters | 0.200 | 0.405 | | | | | |
| born in Germany | 0.900 | 0.304 | | | | | |
| migration background | 0.200 | 0.405 | | | | | |

*Notes:* Main experiment was conducted in December 2015, descriptive statistics for 120 participants are presented in Panel A. Panel B (Beauty rating) was carried out in March 2016 with 40 participants. CRT refers to "Cognitive Reflection Test" proposed by Frederick (2005), SD = standard deviation.

Table A5.2: The effect of beauty, linear controls for CV's characteristics.

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| beauty | 0.024*** | 0.015 | 0.031*** | 0.005 | 0.040*** |
| | (0.008) | (0.013) | (0.010) | (0.009) | (0.011) |
| $2^{nd}$ position | 0.015 | 0.051 | -0.021 | -0.004 | 0.038 |
| | (0.022) | (0.036) | (0.035) | (0.032) | (0.034) |
| $3^{rd}$ position | 0.025 | 0.038 | 0.006 | -0.012 | 0.064* |
| | (0.028) | (0.038) | (0.038) | (0.038) | (0.033) |
| $4^{th}$ position | 0.041* | 0.086** | -0.008 | -0.001 | 0.076*** |
| | (0.021) | (0.036) | (0.031) | (0.025) | (0.027) |
| quality | 0.115*** | 0.112*** | 0.119*** | 0.171*** | 0.064*** |
| | (0.014) | (0.013) | (0.021) | (0.019) | (0.015) |
| experience | 0.045*** | 0.029*** | 0.059*** | 0.048*** | 0.049*** |
| | (0.010) | (0.009) | (0.017) | (0.012) | (0.013) |
| education | 0.161*** | 0.150*** | 0.170*** | 0.130*** | 0.196*** |
| | (0.013) | (0.014) | (0.021) | (0.019) | (0.016) |
| English | 0.021** | -0.012 | 0.052*** | 0.014 | 0.034** |
| | (0.010) | (0.012) | (0.017) | (0.012) | (0.015) |
| unemployed | -0.003 | -0.022* | 0.015 | 0.007 | -0.006 |
| | (0.011) | (0.013) | (0.016) | (0.014) | (0.015) |
| wage | 0.003 | 0.005 | -0.000 | 0.016 | -0.004 |
| | (0.009) | (0.008) | (0.018) | (0.012) | (0.013) |
| computer | 0.041*** | 0.054*** | 0.030** | 0.026** | 0.064*** |
| | (0.010) | (0.013) | (0.014) | (0.012) | (0.013) |
| Constant | 0.097 | 0.151*** | 0.053 | 0.067 | 0.068 |
| | (0.071) | (0.053) | (0.131) | (0.069) | (0.098) |
| Observations | 4,384 | 2,188 | 2,196 | 2,176 | 2,208 |
| Adj. $R^2$ | 0.078 | 0.086 | 0.085 | 0.079 | 0.109 |

*Notes:* This table shows the relationship between applicant's appearance and position of the CV and the chances that the CV is selected for a job interview. *beauty* is a double-standardised beauty score of the photo. *position* refers to the order of appearance of a CV within the occupation (with the first applicant as the reference group). The variables *quality*, *experience*, *education*, *english*, *unemployed*, *wage*, and *computer* refer to the characteristics of each CV (see page 90 for a detailed discussion of the variables). They are rescaled so that a higher value is always better for the employer. Robust two-way clustered (by photo and participant) standard errors in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Table A5.3: The effect of beauty & wingman, two way clustered

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| beauty | 0.060*** | 0.058*** | 0.061*** | 0.040* | 0.076*** |
| | (0.015) | (0.018) | (0.023) | (0.022) | (0.019) |
| wingman beauty | -0.041* | -0.018 | -0.069* | -0.020 | -0.061** |
| | (0.023) | (0.024) | (0.036) | (0.027) | (0.029) |
| beauty × samegender | -0.039* | -0.039 | -0.035 | -0.024 | -0.052** |
| | (0.020) | (0.028) | (0.024) | (0.024) | (0.025) |
| samegender | 0.015 | 0.014 | 0.015 | -0.010 | 0.034 |
| | (0.018) | (0.027) | (0.022) | (0.025) | (0.024) |
| photo female | -0.031 | -0.067*** | 0.006 | -0.039 | -0.018 |
| | (0.019) | (0.023) | (0.035) | (0.036) | (0.026) |
| Turkish origin | 0.048 | 0.015 | 0.087* | 0.096* | -0.013 |
| | (0.035) | (0.042) | (0.050) | (0.058) | (0.035) |
| headscarf | 0.016 | 0.078 | -0.054 | -0.033 | 0.076 |
| | (0.041) | (0.056) | (0.040) | (0.043) | (0.047) |
| $2^{nd}$ position | 0.012 | 0.046 | -0.023 | -0.008 | 0.035 |
| | (0.023) | (0.036) | (0.037) | (0.033) | (0.034) |
| $3^{rd}$ position | 0.018 | 0.033 | 0.003 | -0.017 | 0.055* |
| | (0.028) | (0.038) | (0.038) | (0.036) | (0.033) |
| $4^{th}$ position | 0.039* | 0.084** | -0.008 | -0.003 | 0.074** |
| | (0.022) | (0.036) | (0.033) | (0.025) | (0.030) |
| time (photo) | 0.098 | 0.069 | 0.078 | -0.637 | 0.428* |
| | (0.314) | (0.419) | (0.367) | (0.479) | (0.259) |
| $(\text{time (photo)})^2$ | 0.300 | 0.155 | 0.634 | 2.999* | -0.818 |
| | (0.987) | (1.290) | (1.307) | (1.641) | (0.628) |
| # samegender applicants | -0.012 | -0.018 | -0.007 | -0.006 | -0.021* |
| | (0.008) | (0.011) | (0.012) | (0.012) | (0.011) |
| Constant | 0.459*** | 0.513*** | 0.409*** | 0.609*** | 0.325*** |
| | (0.038) | (0.068) | (0.069) | (0.054) | (0.058) |
| Characteristics | Categorical | Categorical | Categorical | Categorical | Categorical |
| Observations | 4,384 | 2,188 | 2,196 | 2,176 | 2,208 |
| Adj. $R^2$ | 0.092 | 0.106 | 0.104 | 0.099 | 0.126 |

*Notes:* This table shows the relationship between applicant's appearance and position of the CV and the chances that the CV is selected for a job interview. *beauty* is a double-standardised beauty score of the photo. *wingman beauty* is the average *beauty*-score of the applicants of the same gender for the same position. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *Turkish origin*, *headscarf* and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *position* refers to the order of appearance of a CV within the occupation (with the first applicant as the reference group). *time(photo)* is the relative time each participant used to inspect the photo. *# male applicants* is the number of male applicants in the job, ranging from 0 to 4. Categorical controls for CV's characteristics included in each model (see page 90 for a detailed discussion of the variables). Robust two-way clustered (by photo and participant) standard errors in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Table A5.4: The effect of beauty & wingman, clustered by participant

| Type of Occupation | All | No Contact | Contact | Low Skill | High Skill |
|---|---|---|---|---|---|
| beauty | 0.060*** | 0.058*** | 0.061*** | 0.040* | 0.076*** |
| | (0.015) | (0.020) | (0.022) | (0.023) | (0.021) |
| wingman beauty | -0.041** | -0.018 | -0.069*** | -0.020 | -0.061*** |
| | (0.016) | (0.023) | (0.021) | (0.023) | (0.023) |
| beauty $\times$ samegender | -0.039*** | -0.039* | -0.035 | -0.024 | -0.052*** |
| | (0.015) | (0.021) | (0.022) | (0.023) | (0.017) |
| samegender | 0.015 | 0.014 | 0.015 | -0.010 | 0.034* |
| | (0.015) | (0.022) | (0.021) | (0.022) | (0.021) |
| photo female | -0.031 | -0.067* | 0.006 | -0.039 | -0.018 |
| | (0.025) | (0.035) | (0.036) | (0.039) | (0.034) |
| Turkish origin | 0.048 | 0.015 | 0.087* | 0.096 | -0.013 |
| | (0.038) | (0.053) | (0.053) | (0.059) | (0.050) |
| headscarf | 0.016 | 0.078* | -0.054 | -0.033 | 0.076 |
| | (0.034) | (0.041) | (0.048) | (0.048) | (0.047) |
| $2^{nd}$ position | 0.012 | 0.046 | -0.023 | -0.008 | 0.035 |
| | (0.024) | (0.033) | (0.036) | (0.034) | (0.032) |
| $3^{rd}$ position | 0.018 | 0.033 | 0.003 | -0.017 | 0.055 |
| | (0.027) | (0.035) | (0.038) | (0.035) | (0.036) |
| $4^{th}$ position | 0.039* | 0.084*** | -0.008 | -0.003 | 0.074** |
| | (0.021) | (0.030) | (0.030) | (0.030) | (0.029) |
| time (photo) | 0.098 | 0.069 | 0.078 | -0.637** | 0.428* |
| | (0.259) | (0.329) | (0.347) | (0.321) | (0.239) |
| $(\text{time (photo)})^2$ | 0.300 | 0.155 | 0.634 | 2.999** | -0.818 |
| | (0.962) | (1.072) | (1.457) | (1.196) | (0.636) |
| # samegender applicants | -0.012 | -0.018* | -0.007 | -0.006 | -0.021* |
| | (0.008) | (0.010) | (0.010) | (0.010) | (0.011) |
| Constant | 0.459*** | 0.513*** | 0.409*** | 0.609*** | 0.325*** |
| | (0.045) | (0.068) | (0.063) | (0.056) | (0.060) |
| Characteristics | Categorical | Categorical | Categorical | Categorical | Categorical |
| Observations | 4,384 | 2,188 | 2,196 | 2,176 | 2,208 |
| Adj. $R^2$ | 0.092 | 0.106 | 0.104 | 0.099 | 0.126 |

*Notes:* This table shows the relationship between applicant's appearance and position of the CV and the chances that the CV is selected for a job interview. *beauty* is a double-standardised beauty score of the photo. *wingman beauty* is the average *beauty*-score of the applicants of the same gender for the same position. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *Turkish origin*, *headscarf*, and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *position* refers to the order of appearance of a CV within the occupation (with the first applicant as the reference group). *time(photo)* is the relative time each participant used to inspect the photo. *# male applicants* is the number of male applicants in the job, ranging from 0 to 4 Categorical controls for CV's characteristics included in each model (see page 90 for a detailed discussion of the variables). Robust clustered (by participant) standard errors in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

# Table A5.5: Separate regressions for $1^{st}$ and $2^{nd}$ choice (male only)

| Type of Occupation | All | | No Contact | | Contact | | Low Skill | | High Skill | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice |
| beauty | 0.021 | 0.034 | 0.000 | 0.044$^{**}$ | 0.034 | 0.021 | 0.009 | 0.003 | 0.036 | 0.066 |
| | (0.016) | (0.025) | (0.018) | (0.021) | (0.025) | (0.066) | (0.026) | (0.035) | (0.027) | (0.043) |
| wingman beauty | -0.032 | -0.063$^{**}$ | -0.022 | -0.016 | -0.041 | -0.115$^{**}$ | -0.023 | -0.027 | -0.050$^{*}$ | -0.093$^{**}$ |
| | (0.024) | (0.030) | (0.031) | (0.036) | (0.031) | (0.046) | (0.032) | (0.043) | (0.029) | (0.039) |
| beauty × samegender | -0.006 | 0.005 | 0.025 | -0.013 | -0.021 | 0.025 | -0.029 | 0.034 | 0.006 | -0.033 |
| | (0.028) | (0.029) | (0.031) | (0.033) | (0.043) | (0.068) | (0.034) | (0.047) | (0.037) | (0.060) |
| samegender | 0.001 | 0.020 | 0.010 | 0.074$^{**}$ | -0.013 | -0.041 | -0.036 | 0.021 | 0.035 | 0.012 |
| | (0.025) | (0.036) | (0.035) | (0.031) | (0.037) | (0.084) | (0.043) | (0.049) | (0.039) | (0.055) |
| photo female | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Turkish origin | -0.032 | 0.022 | -0.072 | 0.038 | -0.008 | -0.003 | -0.004 | 0.064 | -0.054 | -0.003 |
| | (0.035) | (0.049) | (0.068) | (0.049) | (0.058) | (0.142) | (0.066) | (0.083) | (0.057) | (0.086) |
| headscarf | 0.029 | -0.021 | 0.043 | 0.019 | 0.025 | -0.060 | 0.004 | -0.078 | 0.064 | 0.027 |
| | (0.035) | (0.048) | (0.060) | (0.048) | (0.043) | (0.066) | (0.054) | (0.062) | (0.046) | (0.055) |
| $2^{nd}$ position | 0.017 | 0.029 | 0.066 | 0.074 | -0.027 | -0.027 | 0.015 | 0.046 | 0.019 | 0.008 |
| | (0.039) | (0.037) | (0.048) | (0.053) | (0.053) | (0.050) | (0.048) | (0.052) | (0.045) | (0.043) |
| $3^{rd}$ position | 0.018 | 0.092$^{**}$ | 0.066 | 0.106$^{*}$ | -0.030 | 0.066 | 0.027 | 0.030 | 0.001 | 0.152$^{***}$ |
| | (0.033) | (0.039) | (0.043) | (0.060) | (0.043) | (0.047) | (0.044) | (0.053) | (0.044) | (0.046) |
| $4^{th}$ position | -0.077$^{**}$ | 0.137$^{***}$ | -0.033 | 0.172$^{***}$ | -0.115$^{***}$ | 0.097$^{*}$ | -0.092$^{**}$ | 0.075$^{**}$ | -0.069 | 0.195$^{***}$ |
| | (0.033) | (0.029) | (0.043) | (0.044) | (0.034) | (0.058) | (0.038) | (0.033) | (0.045) | (0.035) |
| 1 year of experience | 0.058$^{**}$ | 0.023 | 0.061 | 0.014 | 0.059$^{**}$ | 0.025 | 0.058 | 0.028 | 0.069$^{**}$ | 0.029 |
| | (0.023) | (0.039) | (0.040) | (0.036) | (0.024) | (0.056) | (0.048) | (0.045) | (0.035) | (0.047) |
| 2 years of experience | 0.085$^{***}$ | 0.031 | 0.073$^{**}$ | -0.035 | 0.101$^{***}$ | 0.080 | 0.092$^{**}$ | 0.073 | 0.084$^{**}$ | 0.006 |
| | (0.030) | (0.036) | (0.036) | (0.040) | (0.036) | (0.061) | (0.043) | (0.053) | (0.042) | (0.044) |
| 3 years of experience | 0.112$^{***}$ | 0.021 | 0.128$^{***}$ | -0.059 | 0.106$^{**}$ | 0.082 | 0.136$^{***}$ | 0.046 | 0.098$^{**}$ | 0.025 |
| | (0.032) | (0.046) | (0.044) | (0.048) | (0.052) | (0.067) | (0.043) | (0.061) | (0.046) | (0.067) |
| asked wage: 20% more than average | -0.098$^{***}$ | -0.035 | -0.112$^{**}$ | -0.034 | -0.102$^{***}$ | -0.018 | -0.172$^{***}$ | -0.070 | -0.024 | -0.008 |
| | (0.035) | (0.045) | (0.054) | (0.054) | (0.032) | (0.074) | (0.051) | (0.058) | (0.041) | (0.055) |
| asked wage: 10% more than average | -0.026 | -0.002 | -0.093$^{**}$ | 0.017 | 0.032 | -0.004 | -0.111$^{***}$ | -0.015 | 0.051 | 0.012 |
| | (0.034) | (0.037) | (0.039) | (0.047) | (0.051) | (0.059) | (0.030) | (0.050) | (0.051) | (0.052) |
| asked wage: 10% less than average | -0.033 | -0.021 | -0.036 | 0.001 | -0.025 | -0.030 | -0.084$^{**}$ | -0.027 | 0.019 | -0.020 |
| | (0.023) | (0.033) | (0.034) | (0.053) | (0.035) | (0.041) | (0.038) | (0.051) | (0.033) | (0.043) |
| asked wage: 20% less than average | -0.030 | -0.047 | -0.006 | -0.071 | -0.044 | -0.018 | -0.037 | -0.052 | -0.024 | -0.052 |
| | (0.025) | (0.039) | (0.041) | (0.051) | (0.037) | (0.067) | (0.040) | (0.055) | (0.036) | (0.052) |
| education: below average | -0.135$^{***}$ | -0.133$^{***}$ | -0.099$^{***}$ | -0.167$^{***}$ | -0.169$^{***}$ | -0.100$^{**}$ | -0.157$^{***}$ | -0.180$^{***}$ | -0.119$^{***}$ | -0.081$^{*}$ |
| | (0.021) | (0.031) | (0.030) | (0.039) | (0.031) | (0.043) | (0.024) | (0.035) | (0.029) | (0.043) |
| education: above average | 0.101$^{***}$ | 0.117$^{***}$ | 0.130$^{***}$ | 0.070 | 0.080$^{**}$ | 0.166$^{***}$ | 0.059$^{*}$ | 0.060 | 0.142$^{***}$ | 0.184$^{***}$ |
| | (0.029) | (0.027) | (0.037) | (0.043) | (0.036) | (0.035) | (0.033) | (0.043) | (0.038) | (0.037) |
| quality: below average | -0.120$^{***}$ | -0.107$^{***}$ | -0.146$^{***}$ | -0.080$^{***}$ | -0.102$^{***}$ | -0.134$^{***}$ | -0.184$^{***}$ | -0.162$^{***}$ | -0.054$^{**}$ | -0.043 |
| | (0.021) | (0.021) | (0.026) | (0.028) | (0.030) | (0.030) | (0.034) | (0.028) | (0.026) | (0.030) |
| quality: above average | 0.060$^{***}$ | 0.047 | 0.077$^{***}$ | 0.018 | 0.045 | 0.080$^{*}$ | 0.098$^{***}$ | 0.065 | 0.035 | 0.055 |
| | (0.021) | (0.029) | (0.024) | (0.030) | (0.035) | (0.043) | (0.027) | (0.043) | (0.031) | (0.036) |
| 13-18 month unemployed | -0.006 | 0.010 | 0.006 | 0.125$^{**}$ | -0.022 | -0.097 | -0.047 | -0.066 | 0.027 | 0.089 |
| | (0.040) | (0.041) | (0.056) | (0.054) | (0.057) | (0.075) | (0.049) | (0.061) | (0.058) | (0.074) |
| 7-12 month unemployed | -0.028 | 0.025 | -0.023 | 0.076 | -0.034 | -0.037 | -0.013 | -0.042 | -0.052 | 0.091 |
| | (0.033) | (0.040) | (0.049) | (0.064) | (0.039) | (0.059) | (0.048) | (0.048) | (0.042) | (0.071) |
| 1-6 month unemployed | -0.019 | 0.003 | 0.026 | 0.052 | -0.062$^{*}$ | -0.065 | -0.052 | -0.021 | 0.006 | 0.027 |
| | (0.026) | (0.034) | (0.036) | (0.045) | (0.035) | (0.045) | (0.034) | (0.053) | (0.034) | (0.051) |
| computer skills: good | 0.015 | 0.032 | 0.022 | 0.067 | 0.006 | -0.010 | 0.007 | 0.018 | 0.025 | 0.046 |
| | (0.024) | (0.033) | (0.037) | (0.043) | (0.025) | (0.037) | (0.035) | (0.040) | (0.034) | (0.051) |
| computer skills: very good | 0.033$^{*}$ | 0.056 | 0.050$^{**}$ | 0.078 | 0.029 | 0.032 | 0.026 | 0.055 | 0.052 | 0.068 |
| | (0.020) | (0.036) | (0.025) | (0.048) | (0.035) | (0.038) | (0.033) | (0.050) | (0.033) | (0.049) |
| English: advanced | 0.066$^{***}$ | 0.007 | 0.062$^{**}$ | -0.039 | 0.085$^{***}$ | 0.039 | 0.025 | -0.024 | 0.098$^{***}$ | 0.041 |
| | (0.021) | (0.027) | (0.026) | (0.044) | (0.032) | (0.039) | (0.039) | (0.035) | (0.027) | (0.032) |
| English: fluently | 0.022 | -0.013 | 0.011 | -0.072 | 0.043 | 0.030 | 0.026 | -0.040 | 0.016 | 0.022 |
| | (0.019) | (0.030) | (0.029) | (0.050) | (0.027) | (0.046) | (0.030) | (0.050) | (0.026) | (0.035) |
| Observations | 2360 | 1770 | 1172 | 879 | 1188 | 891 | 1176 | 882 | 1184 | 888 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview as $1^{st}$ *Choice* or $2^{nd}$ *Choice* respectively. The number of observations for $2^{nd}$ *Choice* decreases because CV's which were selected as first preference are excluded from these estimations. *Turkish origin*, *headscarf*, and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardised beauty score of the photo. The variables *quality (base: average)*, *experience (base: 0 years of experience)*, *education (base: average)*, *english (base: basic)*, *unemployed (base: currently not unemployed)*, *asked wage (base: average wage)*, and *computer (base: sufficient)* refer to the characteristics of each CV. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearance of a CV within the occupation. Robust standard errors two-way clustered (by photo & participant) are in parentheses. All regressions are estimated using LPM (Linear Probability Model). $^{***}$, $^{**}$ and $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

# Table A5.6: Separate regressions for $1^{st}$ and $2^{nd}$ choice (female only)

| Type of Occupation | All | | No Contact | | Contact | | Low Skill | | High Skill | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice | $1^{st}$ Choice | $2^{nd}$ Choice |
| beauty | 0.031 | 0.032$^{**}$ | 0.016 | 0.063$^{**}$ | 0.049$^{*}$ | 0.002 | 0.033 | 0.059$^{*}$ | 0.028 | 0.015 |
| | (0.023) | (0.014) | (0.024) | (0.026) | (0.026) | (0.022) | (0.032) | (0.035) | (0.045) | (0.030) |
| wingman beauty | -0.006 | -0.019 | 0.017 | -0.022 | -0.029 | -0.027 | 0.003 | -0.024 | -0.012 | -0.022 |
| | (0.014) | (0.030) | (0.028) | (0.048) | (0.038) | (0.035) | (0.023) | (0.038) | (0.025) | (0.041) |
| beauty $\times$ samegender | 0.010 | -0.013 | 0.002 | -0.047 | 0.010 | 0.029 | 0.011 | -0.072 | 0.011 | 0.033 |
| | (0.028) | (0.034) | (0.042) | (0.053) | (0.029) | (0.037) | (0.037) | (0.055) | (0.048) | (0.043) |
| samegender | -0.041$^{*}$ | -0.040 | -0.036 | -0.069 | -0.044 | -0.021 | -0.033 | -0.065 | -0.048 | -0.025 |
| | (0.022) | (0.038) | (0.031) | (0.051) | (0.037) | (0.039) | (0.037) | (0.061) | (0.042) | (0.053) |
| photo female | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Turkish origin | 0.153$^{***}$ | 0.036 | 0.050 | -0.007 | 0.256$^{***}$ | 0.119 | 0.197$^{***}$ | -0.007 | 0.101$^{**}$ | 0.082 |
| | (0.032) | (0.074) | (0.067) | (0.084) | (0.032) | (0.087) | (0.061) | (0.103) | (0.048) | (0.096) |
| headscarf | -0.074$^{**}$ | 0.067 | -0.041 | 0.162 | -0.136$^{***}$ | -0.068 | -0.093 | 0.034 | -0.047 | 0.084 |
| | (0.032) | (0.082) | (0.058) | (0.130) | (0.037) | (0.073) | (0.066) | (0.091) | (0.037) | (0.092) |
| $2^{nd}$ position | -0.006 | -0.007 | 0.027 | -0.045 | -0.034 | 0.030 | -0.084$^{**}$ | -0.019 | 0.071 | 0.008 |
| | (0.030) | (0.035) | (0.038) | (0.047) | (0.044) | (0.054) | (0.041) | (0.049) | (0.050) | (0.058) |
| $3^{rd}$ position | -0.030 | -0.040 | -0.015 | -0.067 | -0.037 | -0.009 | -0.074 | -0.038 | 0.017 | -0.030 |
| | (0.032) | (0.040) | (0.045) | (0.063) | (0.041) | (0.058) | (0.049) | (0.047) | (0.047) | (0.056) |
| $4^{th}$ position | -0.106$^{***}$ | 0.106$^{***}$ | -0.092$^{**}$ | 0.136$^{**}$ | -0.114$^{**}$ | 0.056 | -0.146$^{***}$ | 0.107$^{**}$ | -0.070 | 0.088 |
| | (0.037) | (0.038) | (0.040) | (0.066) | (0.050) | (0.048) | (0.045) | (0.050) | (0.048) | (0.062) |
| 1 year of experience | 0.112$^{***}$ | 0.065$^{*}$ | 0.090$^{**}$ | 0.094$^{*}$ | 0.122$^{***}$ | 0.038 | 0.099$^{***}$ | 0.041 | 0.133$^{***}$ | 0.108$^{*}$ |
| | (0.023) | (0.036) | (0.040) | (0.055) | (0.025) | (0.034) | (0.032) | (0.057) | (0.041) | (0.061) |
| 2 years of experience | 0.127$^{***}$ | 0.124$^{***}$ | 0.110$^{**}$ | 0.103$^{**}$ | 0.156$^{***}$ | 0.156$^{**}$ | 0.109$^{**}$ | 0.083 | 0.149$^{***}$ | 0.178$^{***}$ |
| | (0.031) | (0.039) | (0.053) | (0.046) | (0.041) | (0.076) | (0.052) | (0.074) | (0.039) | (0.058) |
| 3 years of experience | 0.144$^{***}$ | 0.063 | 0.131$^{**}$ | 0.058 | 0.152$^{***}$ | 0.077 | 0.122$^{**}$ | 0.092 | 0.205$^{***}$ | 0.094 |
| | (0.040) | (0.058) | (0.061) | (0.079) | (0.041) | (0.083) | (0.061) | (0.083) | (0.051) | (0.089) |
| asked wage: 20% more than average | 0.034 | -0.018 | -0.005 | -0.023 | 0.077$^{*}$ | -0.018 | 0.029 | -0.061 | 0.020 | 0.017 |
| | (0.040) | (0.046) | (0.061) | (0.072) | (0.046) | (0.074) | (0.062) | (0.068) | (0.057) | (0.076) |
| asked wage: 10% more than average | 0.008 | -0.070$^{*}$ | -0.039 | -0.134$^{**}$ | 0.060 | -0.007 | -0.015 | -0.093$^{**}$ | 0.020 | -0.046 |
| | (0.029) | (0.037) | (0.043) | (0.059) | (0.043) | (0.064) | (0.035) | (0.046) | (0.051) | (0.063) |
| asked wage: 10% less than average | -0.041 | -0.017 | -0.088$^{**}$ | -0.038 | -0.001 | 0.001 | -0.028 | 0.008 | -0.044 | -0.027 |
| | (0.030) | (0.037) | (0.045) | (0.038) | (0.035) | (0.057) | (0.051) | (0.051) | (0.043) | (0.053) |
| asked wage: 20% less than average | -0.070$^{**}$ | 0.005 | -0.121$^{**}$ | -0.020 | -0.022 | 0.029 | -0.093$^{*}$ | -0.005 | -0.026 | 0.038 |
| | (0.031) | (0.044) | (0.055) | (0.066) | (0.028) | (0.071) | (0.056) | (0.062) | (0.045) | (0.057) |
| education: below average | -0.110$^{***}$ | -0.109$^{***}$ | -0.105$^{***}$ | -0.154$^{***}$ | -0.117$^{***}$ | -0.057 | -0.053$^{*}$ | -0.073 | -0.177$^{***}$ | -0.153$^{***}$ |
| | (0.024) | (0.041) | (0.025) | (0.049) | (0.036) | (0.054) | (0.030) | (0.062) | (0.033) | (0.046) |
| education: above average | 0.075$^{***}$ | 0.101$^{***}$ | 0.048 | 0.059 | 0.101$^{**}$ | 0.157$^{***}$ | 0.065 | 0.079$^{*}$ | 0.101$^{***}$ | 0.129$^{***}$ |
| | (0.027) | (0.027) | (0.035) | (0.039) | (0.041) | (0.047) | (0.043) | (0.040) | (0.037) | (0.050) |
| quality: below average | -0.118$^{***}$ | -0.151$^{***}$ | -0.090$^{***}$ | -0.164$^{***}$ | -0.133$^{***}$ | -0.141$^{**}$ | -0.174$^{***}$ | -0.207$^{***}$ | -0.065$^{***}$ | -0.105$^{**}$ |
| | (0.018) | (0.033) | (0.029) | (0.041) | (0.029) | (0.059) | (0.030) | (0.051) | (0.024) | (0.047) |
| quality: above average | 0.042 | -0.024 | 0.077$^{*}$ | -0.043 | 0.007 | -0.014 | 0.055 | 0.008 | 0.037 | -0.031 |
| | (0.028) | (0.035) | (0.040) | (0.050) | (0.030) | (0.056) | (0.044) | (0.053) | (0.032) | (0.043) |
| 13-18 month unemployed | 0.046 | 0.029 | 0.088 | 0.025 | 0.003 | 0.033 | 0.122$^{*}$ | 0.004 | -0.060 | 0.032 |
| | (0.044) | (0.059) | (0.067) | (0.060) | (0.056) | (0.091) | (0.064) | (0.059) | (0.060) | (0.091) |
| 7-12 month unemployed | 0.004 | 0.028 | 0.038 | 0.042 | -0.029 | 0.003 | 0.034 | -0.030 | -0.043 | 0.068 |
| | (0.043) | (0.033) | (0.069) | (0.051) | (0.042) | (0.066) | (0.057) | (0.050) | (0.053) | (0.057) |
| 1-6 month unemployed | -0.017 | 0.004 | 0.001 | 0.005 | -0.017 | 0.011 | -0.003 | 0.005 | -0.034 | -0.000 |
| | (0.039) | (0.033) | (0.045) | (0.048) | (0.049) | (0.061) | (0.054) | (0.042) | (0.050) | (0.050) |
| computer skills: good | 0.040 | 0.040 | 0.076$^{**}$ | 0.068 | 0.003 | 0.009 | 0.055 | 0.014 | 0.041 | 0.076$^{*}$ |
| | (0.033) | (0.033) | (0.037) | (0.044) | (0.044) | (0.043) | (0.042) | (0.048) | (0.030) | (0.042) |
| computer skills: very good | 0.060$^{**}$ | 0.034 | 0.081$^{**}$ | 0.051 | 0.051 | 0.013 | 0.013 | 0.034 | 0.127$^{***}$ | 0.066 |
| | (0.026) | (0.028) | (0.035) | (0.033) | (0.038) | (0.049) | (0.031) | (0.043) | (0.032) | (0.045) |
| English: advanced | 0.022 | 0.027 | -0.051 | 0.001 | 0.104$^{***}$ | 0.068$^{*}$ | 0.010 | 0.023 | 0.035 | 0.033 |
| | (0.026) | (0.030) | (0.032) | (0.048) | (0.036) | (0.041) | (0.037) | (0.049) | (0.035) | (0.038) |
| English: fluently | -0.017 | 0.059$^{**}$ | -0.094$^{**}$ | 0.017 | 0.055$^{*}$ | 0.109$^{**}$ | -0.027 | 0.076$^{*}$ | 0.010 | 0.070$^{*}$ |
| | (0.026) | (0.030) | (0.039) | (0.036) | (0.032) | (0.050) | (0.030) | (0.046) | (0.040) | (0.041) |
| Observations | 2024 | 1518 | 1016 | 762 | 1008 | 756 | 1000 | 750 | 1024 | 768 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview as $1^{st}$ *Choice* or $2^{nd}$ *Choice* respectively. The number of observations for $2^{nd}$ *Choice* decreases because CV's which were selected as first preference are excluded from these estimations. *Turkish origin*, *headscarf*, and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardised beauty score of the photo. The variables *quality (base: average)*, *experience (base: 0 years of experience)*, *education (base: average)*, *english (base: basic)*, *unemployed (base: currently not unemployed)*, *asked wage (base: average wage)*, and *computer (base: sufficient)* refer to the characteristics of each CV. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearance of a CV within the occupation. Robust standard errors two-way clustered (by photo & participant) are in parentheses. All regressions are estimated using LPM (Linear Probability Model). $^{***}$, $^{**}$ and $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

## Table A5.7: Separate regressions for low and high CRT

| Type of Occupation | All | | No Contact | | Contact | | Low Skill | | High Skill | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Low CRT | High CRT | Low CRT | High CRT | Low CRT | High CRT | Low CRT | High CRT | Low CRT | High CRT |
| beauty | 0.046$^{**}$ | 0.074$^{***}$ | 0.066$^{**}$ | 0.050$^{*}$ | 0.027 | 0.096$^{***}$ | 0.042 | 0.035 | 0.043 | 0.106$^{***}$ |
| | (0.022) | (0.020) | (0.028) | (0.026) | (0.031) | (0.029) | (0.029) | (0.033) | (0.028) | (0.020) |
| wingman beauty | -0.030 | -0.057$^{**}$ | -0.010 | -0.026 | -0.061 | -0.082$^{**}$ | -0.031 | -0.013 | -0.028 | -0.109$^{***}$ |
| | (0.031) | (0.028) | (0.035) | (0.031) | (0.044) | (0.041) | (0.035) | (0.045) | (0.040) | (0.031) |
| beauty × samegender | -0.026 | -0.051$^{*}$ | -0.031 | -0.046 | -0.028 | -0.048 | -0.009 | -0.039 | -0.045 | -0.068$^{***}$ |
| | (0.027) | (0.026) | (0.031) | (0.039) | (0.034) | (0.030) | (0.032) | (0.041) | (0.039) | (0.022) |
| samegender | 0.009 | 0.030 | 0.004 | 0.033 | 0.013 | 0.024 | -0.014 | -0.000 | 0.020 | 0.063$^{**}$ |
| | (0.026) | (0.029) | (0.035) | (0.038) | (0.035) | (0.035) | (0.030) | (0.048) | (0.038) | (0.027) |
| photo female | -0.052$^{*}$ | -0.008 | -0.109$^{***}$ | -0.023 | 0.015 | 0.004 | -0.071$^{*}$ | -0.000 | -0.021 | -0.004 |
| | (0.027) | (0.031) | (0.036) | (0.041) | (0.045) | (0.053) | (0.042) | (0.058) | (0.028) | (0.035) |
| Turkish origin | 0.004 | 0.095$^{**}$ | 0.031 | 0.004 | -0.023 | 0.183$^{***}$ | 0.072 | 0.123 | -0.098$^{**}$ | 0.057 |
| | (0.046) | (0.037) | (0.054) | (0.071) | (0.077) | (0.051) | (0.060) | (0.086) | (0.050) | (0.037) |
| headscarf | 0.051 | -0.022 | 0.080 | 0.069 | 0.009 | -0.115$^{*}$ | 0.063 | -0.123$^{**}$ | 0.079 | 0.079 |
| | (0.048) | (0.054) | (0.057) | (0.061) | (0.060) | (0.061) | (0.039) | (0.062) | (0.060) | (0.087) |
| $2^{nd}$ position | 0.003 | 0.032 | 0.042 | 0.068 | -0.043 | 0.009 | 0.028 | -0.030 | -0.022 | 0.112$^{**}$ |
| | (0.034) | (0.037) | (0.042) | (0.058) | (0.055) | (0.053) | (0.052) | (0.041) | (0.048) | (0.054) |
| $3^{rd}$ position | -0.020 | 0.071$^{**}$ | -0.001 | 0.089$^{*}$ | -0.048 | 0.067 | -0.014 | -0.004 | -0.023 | 0.148$^{***}$ |
| | (0.044) | (0.034) | (0.061) | (0.049) | (0.054) | (0.058) | (0.054) | (0.042) | (0.051) | (0.043) |
| $4^{th}$ position | 0.045 | 0.035 | 0.111$^{**}$ | 0.064 | -0.025 | 0.017 | 0.028 | -0.031 | 0.057 | 0.103$^{**}$ |
| | (0.030) | (0.031) | (0.053) | (0.039) | (0.039) | (0.051) | (0.039) | (0.044) | (0.048) | (0.041) |
| 1 year of experience | 0.081$^{***}$ | 0.088$^{***}$ | 0.062 | 0.104$^{**}$ | 0.098$^{***}$ | 0.057 | 0.025 | 0.135$^{***}$ | 0.151$^{***}$ | 0.046 |
| | (0.028) | (0.030) | (0.046) | (0.042) | (0.035) | (0.050) | (0.042) | (0.046) | (0.046) | (0.058) |
| 2 years of experience | 0.132$^{***}$ | 0.122$^{***}$ | 0.083$^{**}$ | 0.081 | 0.188$^{***}$ | 0.146$^{**}$ | 0.082$^{*}$ | 0.195$^{***}$ | 0.178$^{***}$ | 0.074 |
| | (0.029) | (0.032) | (0.038) | (0.050) | (0.063) | (0.064) | (0.049) | (0.047) | (0.048) | (0.054) |
| 3 years of experience | 0.105$^{**}$ | 0.136$^{**}$ | 0.053 | 0.121$^{*}$ | 0.155$^{***}$ | 0.133$^{*}$ | 0.089 | 0.209$^{***}$ | 0.158$^{***}$ | 0.102 |
| | (0.042) | (0.053) | (0.065) | (0.072) | (0.059) | (0.078) | (0.061) | (0.064) | (0.057) | (0.075) |
| asked wage: 20% more than average | -0.047 | -0.039 | -0.083 | -0.053 | -0.025 | -0.019 | -0.095 | -0.123$^{**}$ | -0.011 | 0.011 |
| | (0.052) | (0.040) | (0.071) | (0.051) | (0.069) | (0.073) | (0.059) | (0.053) | (0.067) | (0.072) |
| asked wage: 10% more than average | -0.027 | -0.039 | -0.086$^{*}$ | -0.091$^{*}$ | 0.039 | 0.013 | -0.055 | -0.113$^{**}$ | -0.009 | 0.019 |
| | (0.032) | (0.035) | (0.047) | (0.049) | (0.044) | (0.058) | (0.043) | (0.050) | (0.053) | (0.062) |
| asked wage: 10% less than average | -0.072$^{**}$ | 0.000 | -0.080$^{**}$ | -0.029 | -0.057 | 0.022 | -0.084 | 0.001 | -0.056 | 0.008 |
| | (0.029) | (0.032) | (0.035) | (0.030) | (0.040) | (0.052) | (0.054) | (0.045) | (0.039) | (0.042) |
| asked wage: 20% less than average | -0.043 | -0.051 | -0.079 | -0.069 | -0.010 | -0.034 | -0.102$^{*}$ | -0.011 | 0.027 | -0.076$^{*}$ |
| | (0.044) | (0.041) | (0.053) | (0.057) | (0.066) | (0.064) | (0.059) | (0.051) | (0.062) | (0.046) |
| education: below average | -0.200$^{***}$ | -0.157$^{***}$ | -0.230$^{***}$ | -0.157$^{***}$ | -0.174$^{***}$ | -0.151$^{***}$ | -0.186$^{***}$ | -0.165$^{***}$ | -0.218$^{***}$ | -0.171$^{***}$ |
| | (0.029) | (0.035) | (0.041) | (0.037) | (0.042) | (0.051) | (0.048) | (0.053) | (0.046) | (0.038) |
| education: above average | 0.089$^{***}$ | 0.182$^{***}$ | 0.036 | 0.168$^{***}$ | 0.145$^{***}$ | 0.196$^{***}$ | 0.036 | 0.142$^{***}$ | 0.145$^{***}$ | 0.220$^{***}$ |
| | (0.032) | (0.028) | (0.034) | (0.043) | (0.049) | (0.036) | (0.041) | (0.043) | (0.040) | (0.037) |
| quality: below average | -0.140$^{***}$ | -0.220$^{***}$ | -0.145$^{***}$ | -0.198$^{***}$ | -0.137$^{***}$ | -0.238$^{***}$ | -0.213$^{***}$ | -0.320$^{***}$ | -0.070$^{**}$ | -0.127$^{***}$ |
| | (0.028) | (0.031) | (0.034) | (0.043) | (0.036) | (0.040) | (0.044) | (0.050) | (0.033) | (0.033) |
| quality: above average | 0.049$^{**}$ | 0.040 | 0.053$^{*}$ | 0.039 | 0.040 | 0.038 | 0.075$^{**}$ | 0.086$^{**}$ | 0.032 | 0.007 |
| | (0.021) | (0.030) | (0.027) | (0.051) | (0.040) | (0.039) | (0.036) | (0.043) | (0.028) | (0.037) |
| 13-18 month unemployed | 0.028 | 0.005 | 0.097 | 0.037 | -0.047 | -0.018 | 0.024 | -0.078 | 0.013 | 0.046 |
| | (0.058) | (0.048) | (0.078) | (0.073) | (0.074) | (0.065) | (0.061) | (0.068) | (0.081) | (0.084) |
| 7-12 month unemployed | 0.012 | 0.008 | 0.074 | 0.029 | -0.047 | -0.006 | -0.018 | -0.043 | 0.031 | 0.020 |
| | (0.047) | (0.034) | (0.068) | (0.071) | (0.055) | (0.046) | (0.058) | (0.056) | (0.064) | (0.067) |
| 1-6 month unemployed | -0.009 | -0.013 | 0.025 | 0.034 | -0.042 | -0.051 | -0.000 | -0.075 | -0.020 | 0.035 |
| | (0.030) | (0.031) | (0.043) | (0.053) | (0.042) | (0.040) | (0.041) | (0.048) | (0.043) | (0.051) |
| computer skills: good | 0.058$^{*}$ | 0.038 | 0.067$^{*}$ | 0.107$^{**}$ | 0.050 | -0.029 | 0.049 | 0.014 | 0.076$^{**}$ | 0.083 |
| | (0.030) | (0.031) | (0.037) | (0.046) | (0.036) | (0.047) | (0.036) | (0.034) | (0.034) | (0.057) |
| computer skills: very good | 0.044 | 0.088$^{***}$ | 0.061$^{*}$ | 0.125$^{***}$ | 0.046 | 0.047 | 0.026 | 0.070$^{*}$ | 0.080$^{**}$ | 0.140$^{***}$ |
| | (0.027) | (0.033) | (0.035) | (0.047) | (0.033) | (0.052) | (0.035) | (0.039) | (0.038) | (0.050) |
| English: advanced | 0.027 | 0.056 | -0.031 | 0.008 | 0.095$^{**}$ | 0.101$^{**}$ | 0.013 | 0.021 | 0.049 | 0.072$^{*}$ |
| | (0.029) | (0.035) | (0.042) | (0.053) | (0.038) | (0.043) | (0.043) | (0.044) | (0.036) | (0.040) |
| English: fluently | 0.037 | 0.024 | -0.012 | -0.048 | 0.080$^{*}$ | 0.098$^{**}$ | 0.003 | 0.055 | 0.088$^{**}$ | -0.000 |
| | (0.035) | (0.033) | (0.043) | (0.054) | (0.044) | (0.047) | (0.052) | (0.046) | (0.045) | (0.044) |
| Observations | 2288 | 2096 | 1148 | 1040 | 1140 | 1056 | 1132 | 1044 | 1156 | 1052 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview. Recruiters are split by their CRT, which refers to "Cognitive Reflection Test" proposed by Frederick (2005). *Turkish origin*, *headscarf*, and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardised beauty score of the photo. The variables *quality (base: average)*, *experience (base: 0 years of experience)*, *education (base: average)*, *english (base: basic)*, *unemployed (base: currently not unemployed)*, *asked wage (base: average wage)*, and *computer (base: sufficient)* refer to the characteristics of each CV. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearance of a CV within the occupation. Robust standard errors two-way clustered (by photo & participant) are in parentheses. All regressions are estimated using LPM (Linear Probability Model). $^{***}$, $^{**}$ and $^{*}$ denote significance at the 1%, 5%, and 10% level, respectively.

Table A5.8: Separate regressions for $1^{st}$ and $2^{nd}$ block

| Type of Occupation | All | | No Contact | | Contact | | Low Skill | | High Skill | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $1^{st}$ Block | $2^{nd}$ Block | $1^{st}$ Block | $2^{nd}$ Block | $1^{st}$ Block | $2^{nd}$ Block | $1^{st}$ Block | $2^{nd}$ Block | $1^{st}$ Block | $2^{nd}$ Block |
| beauty | 0.077*** | 0.044** | 0.086*** | 0.035 | 0.066** | 0.051 | 0.064* | 0.016 | 0.094*** | 0.059** |
| | (0.019) | (0.021) | (0.019) | (0.029) | (0.029) | (0.032) | (0.035) | (0.024) | (0.025) | (0.028) |
| wingman beauty | -0.066** | -0.024 | -0.029 | -0.011 | -0.107*** | -0.038 | -0.046 | -0.005 | -0.082** | -0.049 |
| | (0.031) | (0.029) | (0.044) | (0.028) | (0.039) | (0.046) | (0.036) | (0.035) | (0.034) | (0.036) |
| beauty × samegender | -0.046* | -0.036 | -0.037 | -0.046 | -0.057 | -0.023 | -0.038 | -0.015 | -0.068* | -0.046* |
| | (0.025) | (0.023) | (0.032) | (0.034) | (0.035) | (0.029) | (0.029) | (0.032) | (0.035) | (0.027) |
| samegender | 0.002 | 0.027 | 0.018 | 0.013 | -0.012 | 0.039 | -0.008 | -0.004 | 0.010 | 0.056* |
| | (0.025) | (0.023) | (0.033) | (0.032) | (0.036) | (0.026) | (0.034) | (0.030) | (0.036) | (0.029) |
| photo female | -0.032 | -0.030 | -0.125*** | -0.019 | 0.075 | -0.040 | -0.086 | 0.006 | 0.013 | -0.049 |
| | (0.041) | (0.027) | (0.036) | (0.032) | (0.064) | (0.046) | (0.078) | (0.023) | (0.030) | (0.043) |
| Turkish origin | 0.056 | 0.033 | 0.048 | -0.018 | 0.083 | 0.083* | 0.173 | 0.034 | -0.057 | 0.021 |
| | (0.062) | (0.037) | (0.046) | (0.060) | (0.102) | (0.046) | (0.109) | (0.036) | (0.039) | (0.060) |
| headscarf | 0.046 | -0.005 | 0.139*** | 0.027 | -0.086 | -0.047 | -0.013 | -0.056 | 0.104*** | 0.051 |
| | (0.043) | (0.046) | (0.040) | (0.065) | (0.070) | (0.042) | (0.060) | (0.041) | (0.031) | (0.058) |
| $2^{nd}$ position | -0.046 | 0.061* | 0.021 | 0.070 | -0.116** | 0.050 | -0.098** | 0.076* | 0.007 | 0.062 |
| | (0.036) | (0.032) | (0.052) | (0.049) | (0.051) | (0.043) | (0.049) | (0.043) | (0.054) | (0.044) |
| $3^{rd}$ position | -0.003 | 0.036 | 0.018 | 0.040 | -0.028 | 0.024 | -0.070 | 0.027 | 0.062 | 0.050 |
| | (0.046) | (0.030) | (0.058) | (0.047) | (0.055) | (0.046) | (0.064) | (0.035) | (0.053) | (0.042) |
| $4^{th}$ position | 0.032 | 0.043* | 0.092* | 0.076* | -0.028 | 0.005 | -0.028 | 0.019 | 0.094* | 0.055 |
| | (0.032) | (0.026) | (0.049) | (0.042) | (0.053) | (0.035) | (0.050) | (0.033) | (0.052) | (0.037) |
| 1 year of experience | 0.099*** | 0.067** | 0.097** | 0.062 | 0.106** | 0.060** | 0.104** | 0.059 | 0.100** | 0.094** |
| | (0.031) | (0.032) | (0.040) | (0.055) | (0.048) | (0.024) | (0.050) | (0.042) | (0.048) | (0.045) |
| 2 years of experience | 0.096*** | 0.153*** | 0.061 | 0.094* | 0.153** | 0.194*** | 0.068 | 0.183*** | 0.130*** | 0.153*** |
| | (0.034) | (0.037) | (0.041) | (0.050) | (0.067) | (0.053) | (0.062) | (0.049) | (0.048) | (0.059) |
| 3 years of experience | 0.077* | 0.153*** | 0.082** | 0.093 | 0.090 | 0.180*** | 0.078 | 0.184*** | 0.093 | 0.184*** |
| | (0.044) | (0.049) | (0.041) | (0.073) | (0.077) | (0.058) | (0.069) | (0.054) | (0.068) | (0.070) |
| asked wage: 20% more than average | -0.049 | -0.040 | -0.073 | -0.069 | -0.058 | -0.002 | -0.072 | -0.144** | -0.053 | 0.035 |
| | (0.035) | (0.048) | (0.055) | (0.064) | (0.065) | (0.062) | (0.070) | (0.058) | (0.057) | (0.064) |
| asked wage: 10% more than average | -0.063 | -0.006 | -0.092 | -0.087** | -0.052 | 0.086 | -0.075* | -0.091** | -0.064 | 0.059 |
| | (0.040) | (0.033) | (0.057) | (0.041) | (0.048) | (0.055) | (0.045) | (0.043) | (0.066) | (0.048) |
| asked wage: 10% less than average | -0.080** | -0.001 | -0.024 | -0.080* | -0.115** | 0.064 | -0.101** | 0.004 | -0.062 | 0.014 |
| | (0.032) | (0.033) | (0.044) | (0.047) | (0.046) | (0.043) | (0.045) | (0.047) | (0.048) | (0.042) |
| asked wage: 20% less than average | -0.082* | -0.015 | -0.081 | -0.068 | -0.064 | 0.018 | -0.070 | -0.047 | -0.082 | 0.040 |
| | (0.044) | (0.042) | (0.052) | (0.059) | (0.066) | (0.057) | (0.057) | (0.050) | (0.066) | (0.053) |
| education: below average | -0.162*** | -0.196*** | -0.181*** | -0.216*** | -0.140*** | -0.179*** | -0.164*** | -0.178*** | -0.161*** | -0.230*** |
| | (0.035) | (0.030) | (0.033) | (0.037) | (0.045) | (0.049) | (0.048) | (0.038) | (0.043) | (0.048) |
| education: above average | 0.115*** | 0.151*** | 0.084** | 0.114*** | 0.174*** | 0.185*** | 0.056 | 0.107*** | 0.171*** | 0.214*** |
| | (0.037) | (0.027) | (0.040) | (0.040) | (0.052) | (0.041) | (0.045) | (0.041) | (0.050) | (0.044) |
| quality: below average | -0.176*** | -0.181*** | -0.224*** | -0.135*** | -0.145*** | -0.212*** | -0.235*** | -0.290*** | -0.109*** | -0.087*** |
| | (0.031) | (0.020) | (0.035) | (0.035) | (0.035) | (0.030) | (0.052) | (0.044) | (0.032) | (0.025) |
| quality: above average | 0.037 | 0.049 | 0.052 | 0.034 | 0.026 | 0.068 | 0.086** | 0.059 | 0.001 | 0.055* |
| | (0.025) | (0.032) | (0.037) | (0.039) | (0.028) | (0.042) | (0.041) | (0.047) | (0.035) | (0.033) |
| 13-18 month unemployed | 0.051 | -0.003 | 0.084 | 0.070 | -0.020 | -0.049 | 0.025 | -0.048 | 0.056 | 0.014 |
| | (0.040) | (0.058) | (0.069) | (0.068) | (0.054) | (0.068) | (0.057) | (0.069) | (0.065) | (0.078) |
| 7-12 month unemployed | 0.039 | -0.015 | 0.039 | 0.060 | 0.009 | -0.073 | 0.022 | -0.066 | 0.042 | 0.012 |
| | (0.031) | (0.047) | (0.066) | (0.056) | (0.046) | (0.063) | (0.045) | (0.059) | (0.046) | (0.062) |
| 1-6 month unemployed | -0.009 | -0.015 | 0.026 | 0.030 | -0.063 | -0.041 | -0.037 | -0.038 | 0.011 | -0.004 |
| | (0.024) | (0.032) | (0.046) | (0.043) | (0.042) | (0.042) | (0.038) | (0.048) | (0.041) | (0.046) |
| computer skills: good | 0.048 | 0.050 | 0.098** | 0.080** | 0.028 | -0.007 | -0.012 | 0.068** | 0.121*** | 0.064 |
| | (0.034) | (0.032) | (0.050) | (0.034) | (0.040) | (0.044) | (0.044) | (0.034) | (0.043) | (0.042) |
| computer skills: very good | 0.060** | 0.070** | 0.099** | 0.082** | 0.064* | 0.026 | 0.018 | 0.059* | 0.124*** | 0.118*** |
| | (0.029) | (0.029) | (0.046) | (0.037) | (0.036) | (0.045) | (0.035) | (0.034) | (0.040) | (0.046) |
| English: advanced | 0.054** | 0.034 | -0.003 | -0.015 | 0.110*** | 0.087** | 0.064 | -0.024 | 0.031 | 0.091** |
| | (0.027) | (0.024) | (0.042) | (0.032) | (0.033) | (0.040) | (0.043) | (0.029) | (0.025) | (0.037) |
| English: fluently | 0.053** | 0.013 | 0.029 | -0.069** | 0.084* | 0.091** | 0.048 | -0.007 | 0.076 | 0.057 |
| | (0.027) | (0.018) | (0.039) | (0.034) | (0.045) | (0.036) | (0.039) | (0.035) | (0.048) | (0.041) |
| Observations | 1920 | 2464 | 960 | 1228 | 960 | 1236 | 960 | 1216 | 960 | 1248 |

*Notes:* This table shows the relationship between applicant's appearance and their characteristics and the chances that a CV is selected for a job interview as $1^{st}$ *Block* or $2^{nd}$ *Block* of the experiment respectively. *Turkish origin*, *headscarf*, and *photo female* are dummy variables if the applicant is ethnic Turkish, wears a headscarf or is female respectively. *beauty* is a double-standardised beauty score of the photo. The variables *quality (base: average)*, *experience (base: 0 years of experience)*, *education (base: average)*, *english (base: basic)*, *unemployed (base: currently not unemployed)*, *asked wage (base: average wage)*, and *computer (base: sufficient)* refer to the characteristics of each CV. *samegender* is a dummy variable which is equal to 1 if the decision maker and the applicant are of the same gender. *position* refers to the order of appearance of a CV within the occupation. Robust standard errors two-way clustered (by photo & participant) are in parentheses. All regressions are estimated using LPM (Linear Probability Model). ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

# Bibliography


Akrami, N., B. Ekehammar, and T. Araya (2000). Classical and modern racial prejudice: A study of attitudes toward immigrants in Sweden. *European Journal of Social Psychology 30*(4), 521–532.

Alesina, A. and N. Fuchs-Schündeln (2007). Goodbye Lenin (or not?): The effect of communism on people's preferences. *American Economic Review 97*(4), 1507–1528.

Almås, I., A. W. Cappelen, K. G. Salvanes, E. Ø. Sørensen, and B. Tungodden (2015). Willingness to compete: Family matters. *Management Science 62*(8), 2149–2162.

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. In *Handbook of Labor Economics*, Volume 3, pp. 3143 – 3259. Elsevier.

Andreoni, J. and L. Vesterlund (2001). Which is the Fair Sex? Gender Differences in Altruism. *The Quarterly Journal of Economics 116*(1), 293 – 312.

Antonovics, K., P. Arcidiacono, and R. Walsh (2009). The effects of gender interactions in the lab and in the field. *The Review of Economics and Statistics 91*(1), 152–162.

Ariely, D. and T. S. Wallsten (1995). Seeking Subjective Dominance in Multidimensional Space: An Explanation of the Asymmetric Dominance Effect. *Organizational Behavior and Human Decision Processes 63*(3), 223 – 232.

Backus, P., M. Cubel, M. Guid, S. Sánchez-Pages, and E. Mañas (2016). Gender, competition and performance: Evidence from real tournaments. Ieb working paper 2016/27.

Baert, S. (2018). Hiring discrimination: an overview of (almost) all correspondence experiments since 2005. In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, pp. 63–77. Springer.

Baert, S., A. Albanese, S. du Gardein, J. Ovaere, and J. Stappers (2017). Does work experience mitigate discrimination? *Economics Letters 155*, 35–38.

Baert, S. and A.-S. De Pauw (2014). Is ethnic discrimination due to distaste or statistics? *Economics Letters 125*(2), 270–273.

Balafoutas, L., R. Kerschbamer, and M. Sutter (2012). Distributional preferences and competitive behavior. *Journal of Economic Behavior & Organization 83*(1), 125 – 135.

Bartling, B., E. Fehr, M. A. Maréchal, and D. Schunk (2009). Egalitarianism and Competitiveness. *The American Economic Review 99*(2), 93 – 98.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016). Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition. *American Economic Review 106*(6), 1437–75.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Bertrand, M. and E. Duflo (2017). Field Experiments on Discrimination. In *Handbook of Economic Field Experiments*, Volume 1, pp. 309–393. Elsevier.

Bertrand, M. and S. Mullainathan (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review 91*(2), 67–72.

Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review 94*(4), 991–1013.

Biddle, J. E. and D. S. Hamermesh (1998). Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre. *Journal of Labor Economics 16*(1), 172–201.

Blau, F. D. and J. DeVaro (2007). New evidence on gender differences in promotion rates: An empirical analysis of a sample of new hires. *Industrial Relations: A Journal of Economy and Society 46*(3), 511–550.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources 8*(4), 436–455.

Böheim, R., C. Freudenthaler, and M. Lackner (2016). Gender differences in risk-taking: Evidence from professional basketball. IZA Discussion Paper 10011, Institute of Labor Economics (IZA).

Bohren, A., A. O. Imas, and M. Rosenberg (forthcoming). The Dynamics of Discrimination: Theory and Evidence. *American Economic Review*.

Boll, C. and A. Lagemann (2018). Gender pay gap in EU countries based on SES (2014). Luxembourg, Publication Office of the European Union.

Booth, A. and P. Nolen (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization 81*(2), 542–555.

Booth, A. and E. Yamamura (2018). Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in Japan. *Review of Economics and Statistics 100*(4), 581–593.

Brenner, L., Y. Rottenstreich, and S. Sood (1999). Comparison, Grouping, and Preference. *Psychological Science 10*(3), 225–229.

Buser, T., N. Peter, and S. Wolter (2017). Gender, Willingness to Compete and Career Choices along the Whole Ability Distribution. IZA Discussion Papers 10976, Institute of Labor Economics (IZA).

Cameron, A. C. and D. L. Miller (2015). A Practitioner's Guide to Cluster-Robust inference. *Journal of Human Resources 50*(2), 317–372.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge university press.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Carlsson, M. and D.-O. Rooth (2007). Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics 14*(4), 716 – 729.

Cipriani, G. P. and A. Zago (2011). Productivity or discrimination? Beauty and the exams. *Oxford Bulletin of Economics and Statistics 73*(3), 428–447.

Cohen-Zada, D., A. Krumer, M. Rosenboim, and O. M. Shapir (2017). Choking under pressure and gender: Evidence from professional tennis. *Journal of Economic Psychology 61*, 176–190.

Corneo, G. (2001). Inequality and the State: Comparing US and German Preferences. *Annales d'Économie et de Statistique 63/64*, 283–296.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–74.

Czibor, E., J. Claussen, and M. Van Praag (2019). Women in a men's world: Risk taking in an online card game community. *Journal of Economic Behavior & Organization 158*, 62–89.

Darwin, C. (1871). *The descent of man and selection in relation to sex*, Volume 1. Murray.

de Sousa, J., G. Hollard, et al. (2015). Gender differences: evidence from field tournaments. Technical report, CEPREMAP.

Derous, E., A. Buijsrogge, N. Roulin, and W. Duyck (2016). Why your stigma isn't hired: A dual-process framework of interview bias. *Human Resource Management Review 26*(2), 90 – 111.

Dittrich, M., A. Knabe, and K. Leipold (2014). Gender differences in experimental wage negotiations. *Economic Inquiry 52*(2), 862–873.

Dohmen, T. and A. Falk (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review 101*(2), 556–90.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005). Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. IZA Discussion Paper 1730, Institute of Labor Economics (IZA).

Dreber, A., E. von Essen, and E. Ranehill (2011). Outrunning the gender gap - boys and girls compete equally. *Experimental Economics 14*(4), 567–582.

Eckel, C. C. and P. J. Grossman (1998). Are Women Less Selfish Than Men?: Evidence From Dictator Experiments. *The Economic Journal 108*(448), 726–735.

Edo, A., N. Jacquemet, and C. Yannelis (2019). Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Review of Economics of the Household 17*(1), 349–376.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics 14*(4), 583 – 610.

Exley, C. L., M. Niederle, and L. Vesterlund (forthcoming). Knowing when to ask: The cost of leaning in. *Journal of Political Economy*.

Fehr, E., D. Glätzle-Rützler, and M. Sutter (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review 64*, 369 – 383.

Fehr, E. and K. M. Schmidt (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics 114*(3), 817 – 868.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *The Journal of Economic Perspectives 19*(4), 25 – 42.

Fryer Jr, R. G. and S. D. Levitt (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics 119*(3), 767–805.

Gebauer, J. E., M. R. Leary, and W. Neberich (2012). Unfortunate first names: Effects of name-based relational devaluation and interpersonal neglect. *Social Psychological and Personality Science 3*(5), 590–596.

Gerdes, C. and P. Gränsmark (2010). Strategic behavior across gender: a comparison of female and male expert chess players. *Labour Economics 17*(5), 766–775.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Gneezy, U. and A. Rustichini (2004). Gender and competition at a young age. *The American Economic Review 94*(2), 377–381.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik 239*(2), 345–360.

Güth, W., C. Schmidt, and M. Sutter (2007). Bargaining outside the lab – a newspaper experiment of a three-person ultimatum game. *The Economic Journal 117*(518), 449 – 469.

Haan, T., T. Offerman, and R. Sloof (2017). Discrimination in the Labour Market: The Curse of Competition between Workers. *The Economic Journal 127*(603), 1433–1466.

Hamermesh, D. and J. Biddle (1994). Beauty and the Labor Market. *American Economic Review 84*(5), 1174–94.

Hamermesh, D. S. (1986). Chapter 8 the demand for labor in the long run. In *Handbook of Labor Economics*, Volume 1, pp. 429 – 471. Elsevier.

Hamermesh, D. S. (2011). *Beauty pays: Why attractive people are more successful*. Princeton University Press.

Hedegaard, M., R. Kerschbamer, and J.-R. Tyran (2011). Correlates and Consequences of Distributional Preferences: An Internet Experiment. Department of Economics, University of Copenhagen, Mimeo.

Hernández-Julián, R. and C. Peters (2018). Physical appearance and peer effects in academic performance. *Applied Economics Letters 25*(13), 887–890.

Huber, J., J. W. Payne, and C. Puto (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research 9*(1), 90–98.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal 8*(4), 453–479.

Jelnov, P. (2018). Latent Dimensions and Similarity. Technical report.

John, K. and S. L. Thomsen (2015). School-track environment or endowment: What determines different other-regarding behavior across peer groups? *Games and Economic Behavior 94*, 122 – 141.

Kaas, L. and C. Manger (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review 13*(1), 1–20.

Kahneman, D. and P. Egan (2011). *Thinking, fast and slow*, Volume 1. Farrar, Straus and Giroux New York.

Kahneman, D. and S. Frederick (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: The psychology of intuitive judgment.*, pp. 49–81. New York, NY, US: Cambridge University Press.

Kerschbamer, R. (2015). The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test. *European Economic Review 76*, 85 – 103.

Kraft, P. (2012). *The Role of Beauty in the Labor Market*. Ph. D. thesis, Dissertation.

Kuhn, A. (2013). Inequality Perceptions, Distributional Norms, and Redistributive Preferences in East and West Germany. *German Economic Review 14*(4), 483–499.

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

Leckcivilize, A. and A. Straub (2018). Headscarf and job recruitment—lifting the veil of labour market discrimination. *IZA Journal of Labor Economics 7*(1), 11.

Lee, S., M. Niederle, and N. Kang (2014). Do single-sex schools make girls more competitive? *Economics Letters 124*(3), 474–477.

Leibbrandt, A. and J. A. List (2015). Do women avoid salary negotiations? Evidence from a large-scale natural field experiment. *Management Science 61*(9), 2016–2024.

Lichters, M., M. Sarstedt, and B. Vogt (2015). On the practical relevance of the attraction effect: A cautionary note and guidelines for context effect experiments. *AMS Review 5*(1), 1–19.

Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being. *The Quarterly Journal of Economics 120*(3), 963–1002.

Manant, M., S. Pajak, and N. Soulié (2019). Can social media lead to labor market discrimination? Evidence from a field experiment. *Journal of Economics & Management Strategy 28*(2), 225–246.

Manning, A. and J. Swaffield (2008). The gender gap in early-career wage growth. *The Economic Journal 118*(530), 983–1024.

Mobius, M. M. and T. S. Rosenblat (2006). Why beauty matters. *American Economic Review 96*(1), 222–235.

Mueller, G. and E. Plug (2006). Estimating the Effect of Personality on Male and Female Earnings. *ILR Review 60*(1), 3–22.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review of Economics 3*(1), 601–630.

Nunley, J. M., A. Pugh, N. Romero, and R. A. Seals (2015). Racial discrimination in the labor market for recent college graduates: Evidence from a field experiment. *The BE Journal of Economic Analysis & Policy 15*(3), 1093–1125.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review 14*(3), 693–709.

Oosterbeek, H., R. Sloof, and G. van de Kuilen (2004). Cultural differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics 7*(2), 171–188.

Oppenheimer, D. M., T. Meyvis, and N. Davidenko (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power . *Journal of Experimental Social Psychology 45*(4), 867 – 872.

Ors, E., F. Palomino, and E. Peyrache (2013). Performance gender gap: does competition matter? *Journal of Labor Economics 31*(3), 443–499.

Pacini, D. and F. Windmeijer (2016). Robust inference for the Two-Sample 2SLS estimator. *Economics Letters 146*, 50 – 54.

Rammstedt, B. and O. P. John (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality 41*(1), 203 – 212.

Reuben, E., P. Sapienza, and L. Zingales (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences 111*(12), 4403–4408.

Rooth, D. O. (2009). Obesity, attractiveness, and differential treatment in hiring a field experiment. *Journal of Human Resources 44*(3), 710–735.

Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics 17*(3), 523–534.

Slaughter, J. E., E. E. Kausel, and M. A. Quiñones (2011). The decoy effect as a covert influence tactic. *Journal of Behavioral Decision Making 24*(3), 249–266.

Valfort, M.-A. (2017). Has France a Problem with Muslims? Evidence from a Field Experiment in the Labour Market. Technical report.

Weichselbaumer, D. (2016). Discrimination against Female Migrants Wearing Headscarves. IZA Discussion Paper 10217, Institute of Labor Economics (IZA).

Weichselbaumer, D. and J. Schuster (2017). The effect of photos and name change on discrimination against migrants in Austria 2017. Technical report.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

Zschirnt, E. and D. Ruedin (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies 42*(7), 1115–1134.