Be a man or become a nurse: comparing gender discrimination by employers across a wide variety of professions

Kübler, Dorothea; Schmid, Julia; Stüber, Robert

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
Wissenschaftszentrum Berlin für Sozialforschung (WZB)

Empfohlene Zitierung / Suggested Citation:
Kübler, D., Schmid, J., & Stüber, R. (2017). *Be a man or become a nurse: comparing gender discrimination by employers across a wide variety of professions.* (Discussion Papers / Wissenschaftszentrum Berlin für Sozialforschung, Forschungsschwerpunkt Markt und Entscheidung, Abteilung Verhalten auf Märkten, SP II 2017-201). Berlin: Wissenschaftszentrum Berlin für Sozialforschung gGmbH. http://hdl.handle.net/10419/157371

Kübler, Dorothea; Schmid, Julia; Stüber, Robert

Working Paper

Be a man or become a nurse: Comparing gender discrimination by employers across a wide variety of professions

WZB Discussion Paper, No. SP II 2017-201

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Kübler, Dorothea; Schmid, Julia; Stüber, Robert (2017) : Be a man or become a nurse: Comparing gender discrimination by employers across a wide variety of professions, WZB Discussion Paper, No. SP II 2017-201, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin

This Version is available at:
http://hdl.handle.net/10419/157371

WZB

Wissenschaftszentrum Berlin
für Sozialforschung

Dorothea Kübler
Julia Schmid
Robert Stüber

Be a Man or Become a Nurse: Comparing Gender Discrimination by Employers across a Wide Variety of Professions

Discussion Paper

SP II 2017–201

April 2017

WZB Berlin Social Science Center

Research Area
Markets and Choice

Research Unit
Market Behavior

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Dorothea Kübler, Julia Schmid, Robert Stüber
Be a Man or Become a Nurse: Comparing Gender Discrimination by Employers across a Wide Variety of Professions

Affiliation of the authors:

Dorothea Kübler
WZB Berlin Social Science Center & TU Berlin

Julia Schmid
DIW

Robert Stüber
BDPEMS & WZB Belin Social Science Center

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Abstract

Be a Man or Become a Nurse: Comparing Gender Discrimination by Employers across a Wide Variety of Professions

by Dorothea Kübler, Julia Schmid and Robert Stüber*

We investigate gender discrimination and its variation between firms, occupations, and industries with a factorial survey design (vignette study) for a large sample of German firms. Short CVs of fictitious applicants are presented to human resource managers who indicate the likelihood of the applicants being invited to the next step of the hiring process. We observe that women are evaluated worse than men on average, controlling for all other attributes of the CV, i.e., school grades, age, information about activities since leaving school, parents' occupations etc. Discrimination against women varies across industries and occupations, and is strongest for occupations with lower educational requirements and of lower occupational status. Women receive worse evaluations when applying for male-dominated occupations. Overall, the share of women in an occupation explains more of the difference in evaluations than any other occupation- or firm-related variable.

Keywords: Gender discrimination, hiring decisions, vignette study

JEL classification: C99; J71

* E-mail: dorothea.kuebler@wzb.eu, jschmid@diw.de, robert.stueber@wzb.eu.
We would like to thank Heike Solga and Paula Protsch for their collaboration and fruitful discussions, the BIBB (Federal Institute of Vocational Education and Training) which made the study possible, and also Katrin Auspurg who helped us with the selection of the vignettes. We also would like to thank Bernd Fitzenberger, Kristina Strohmaier and Marina Töpfer for their comments. We are grateful to Thu-Ha Nguyen, Michaela Zwiebel, Sajoscha Engelhardt, and Manuela Ludwig for their assistance. The paper has benefited from seminar participants at the Humboldt University of Berlin. Financial support from CRC TR190 is gratefully acknowledged. All remaining errors are our own.

1 Introduction

Labor market outcomes for men and women differ along many dimensions, such as wages earned and the number of working hours. While different trajectories in the labor market may be due to differences unfolding in the life course, such as asymmetries in the effects of having children (see Kleven et al., 2017), we focus on the first stage of people's careers, namely their entry into the labor market. We ask how young women and men fare right after leaving school when applying for their first job. Since girls, on average, do at least as well in school as boys, it can be expected that their chances of getting a job after school should be at least as good as those of young men. We ask whether this is the case by looking at the employment probabilities of men and women early in the career.

We aim to make two contributions, one substantive on discrimination and one methodological. First, our study investigates whether firms from many different industries discriminate by gender when hiring in the entry-level labor market. By gender discrimination we mean that men and women who are equal with respect to all observable productivity-related characteristics are treated differently by employers.[1] While there are many studies that leave no doubt about the existence of discrimination in certain professions, a more comprehensive picture of the labor market is missing in order to establish in which segments of the labor market and for which professions discrimination is observed.

Regarding the methodology, we conducted a survey experiment based on vignettes consisting of the CVs of fictitious applicants. The respondents of the survey were asked to evaluate the fictitious applicants as if they had applied for an apprenticeship position in the firm. This method, which to our knowledge has not been used to investigate labor market discrimination, has a number of advantages, and it complements the use of observational data and field experiments. First, we are able to estimate the causal effect of being female on our outcome variable (the evaluations). By varying the different dimensions of the vignette (items on the CV), we can analyze how gender discrimination varies with the quality of an application. This, together with our outcome variable being non-binary, allows us to address Heckman and Siegelman's critique of field experiments that do not vary the quality of the applicant. They point out that if the variance of unobserved productivity characteristics differs between men and women, taking a snapshot at one quality level of applicants (as most field experiments do) can generate a biased measure of discrimination. Moreover, our survey was conducted with a large sample of German firms hiring apprentices in 126 different professions, which permits us to investigate the firm- and occupation-specific determinants of gender discrimination, using an experimental design.

We focus on the market for apprenticeships, which is the main entry-level labor market for young Germans below the level of tertiary education. More than 60% of all school-leavers start an

[1] This encompasses taste-based and statistical discrimination. Note that statistical discrimination is seen as acceptable in some contexts (young drivers have to pay more for car insurance, for example), while it is considered as unacceptable or illegal in other contexts, such as gender- or race-based hiring decisions.

apprenticeship, and more than half of the apprentices remain employed by the firm that trained them.[2] The hiring procedures for apprentices are similar to those for other employees. The market for apprenticeships is competitive in that a number of applicants do not get a job in their desired profession, and many applicants remain without a job every year. At the same time many slots remain unfilled due to strong regional and occupational differences.

We find economically substantial and statistically significant discrimination against women. The penalty for being female is as large as the effect of having a grade point average that is worse by one grade.[3] In contrast, we do not observe any discrimination based on an applicant's socioeconomic background. Furthermore, the amount of discrimination varies by industry. In line with prior research, it also matters whether an occupation is predominantly female or male, with women being significantly less attractive than men when applying for male-dominated occupations. We do not find evidence of discrimination against men in female-dominated occupations though. We also find no evidence of discrimination if the labor market is very tight. On the other hand, the size of the firm (as measured by the number of employees or the number of apprentices) and the degree of professionalization of the recruitment process do not matter as moderators of gender discrimination.

Prior research suggests that women experience discrimination in high-status occupations. We consider a number of variables that are indicative of the status of a profession. First, we study whether the average salary of a profession moderates discrimination. Then we consider the typical school-leaving qualification required for the profession. Clearly, the average salary of a profession and its required education level may influence gender discrimination for reasons other than status concerns, but both are correlated with indices of occupational status developed in the literature. We finally investigate how these indices of occupational status influence discrimination. Women fare especially worse than men in professions requiring the lowest school degree and for professions with a low social status, according to the status indices. The average wage of a profession has no systematic impact on gender differences in evaluations.

After providing evidence on the correlation between the occupations and industry characteristics with gender discrimination, we determine which of the variables can explain discrimination. The experimental literature has studied the role of the share of women in a profession, but the existing studies have looked at a limited number of professions and are unable to control for the variation of the status of a profession (for example, secretaries as typically female and engineers as typically male occupations). We are able to disentangle the different characteristics of the professions and provide evidence of the ceteris-paribus effect of each of the potential moderator variables on gender discrimination.

Our main finding is that the share of women in a profession explains most of the difference in evaluations. All other firm- and occupation-specific variables cannot explain the discrimination

[2]See BIBB (2012) and BMBF (2012), respectively.

[3]German school grades range from 1 (best) to 6 (worst) where 4 is the passing grade.

observed. These results are novel because they are derived from a large sample of firms, allowing us to control for many possible moderators.

2 Literature and research questions

This study relates to the literature on discrimination in the labor market as well as to methodological contributions regarding the identification of discrimination. We discuss these two strands of the literature in turn and explain how our study contributes to them.

Discrimination in the labor market

Traditionally, labor economists have applied regression-based methods to observational data in order to measure (gender) discrimination in the labor market (see Altonji and Blank, 1999, for an overview). Analyzing both the difference in wages and participation rates, there is evidence of gender discrimination in the labor market. However, this literature suffers from a number of drawbacks (Azmat and Petrongolo, 2014). One important problem is the possibility that discrimination is overestimated, since it is the residual after controlling for observable differences in productivity where these productivity differences might be insufficiently measured. For this reason, many researchers have conducted field experiments to study discrimination in the labor market.[4]

In the experiments, call-back rates, invitations to job interviews or job offers are compared between artificial applicants that are similar with respect to their characteristics but differ, e.g., with respect to race, ethnicity, or sex. Hence, applicants are equal with respect to their observable productivity and differences in outcomes are interpreted as reflecting discrimination.[5] This literature found evidence consistent with labor market discrimination against black people in the United States (Bertrand and Mullainathan, 2004), against Turkish immigrants in Germany (Kaas and Manger, 2011), and against young women for high-skilled jobs in France (Petit, 2007), among many other groups.

Gender discrimination may be driven by existing differences in the share of women and men in an occupation. One possible reason is that the prevalence of one gender causes stereotyping, resulting in a more favorable outcome for the dominant gender in the occupation (Booth and Leigh, 2010). Relatedly, psychological evidence suggests that individuals working in a job where one gender is prevalent think that success in this job requires characteristics typical of that gender (Schein, 1973). This influence of the gender composition of an occupation on discrimination

[4] Alternative approaches based on observational data were also developed, see e.g., Bayard et al. (2003). There are also laboratory studies on the effect of gender on hiring recommendations. In these studies students' or recruiters' are asked to rate the employment suitability of applicants, while applicant gender is varied. These studies find mixed evidence but Olian et al. (1988) conclude from their meta-analysis that there is only little evidence of gender discrimination.

[5] These field studies take on two different forms. In the case of audit studies, applicants trained to act alike apply for jobs, while correspondence studies use written or online applications.

has received some attention in the literature. Table 1 shows studies that compare occupations with a varying share of female employees. Levinson (1975) and Riach and Rich (2006) both find discrimination against men in female-dominated occupations and discrimination against women in male-dominated occupations while in both studies the discrimination against men in female-dominated occupations is more pronounced. Hence, gender discrimination seems to differ across jobs with different shares of female employees. This is especially interesting since the two studies differ with respect to method, location, time, and occupations considered. In addition, in their correspondence study conducted in the UK, Riach and Rich (2006) also find discrimination against males for accountants and computer programmers – two occupations that they classify as gender-neutral. These findings contrast with the results of Riach and Rich (1987) based on a correspondence study conducted in Australia where discrimination against women is found in some (computer programmer and gardener) but not all of the male-dominated occupations. No discrimination against men is found in the female-dominated profession of clerical worker.

Motivated by these studies, Booth and Leigh (2010) consider female-dominated occupations with a varying gender ratio. They find significant discrimination against men for the entire sample and also for the two occupations with a larger share of women (data-entry and waitstaff), but no significant discrimination for the two occupations with a lower share of women (customer service and sales). Note that these two latter occupations are still female-dominated, only to a lower degree. Weichselbaumer (2004) compares interview invitations between a feminine woman, a more masculine woman, and a man in Austria. The women's gender identity is conveyed via different items such as the hobbies and the picture in the CV. Weichselbaumer finds discrimination against both types of women for one of the two occupations that are male-dominated, but not for the other. She also finds discrimination against men (and in favor of masculine and feminine women) for the strongly female-dominated occupation of a secretarial assistant. No statistically significant differences between the three applicant types are found for the female-dominated occupation of an accountant.

Given this evidence, there seems to be a relationship between gender composition and discrimination that deserves closer attention.[6] Our sample allows us to compare occupations with a share of female employers between almost 0 and 100 percent, thereby providing evidence over the entire range of gender ratios. Moreover, we employ two different measures of the gender ratio. Like most previous studies, we base one of our measures on the national average for each apprenticeship occupation. Our second measure of gender dominance is the respondents' self-reported share of female employees in the apprenticeship occupation in the firm.

It has been argued that discrimination is stronger in occupations of higher status and in jobs that are more senior (Riach and Rich, 2002; Azmat and Petrongolo, 2004). This claim is based, in part, on the findings for occupations with different shares of female employees. Also, the

[6]Early evidence from psychology on the influence of an occupation's gender-type on discrimination is mixed (e.g., Cohen and Bunker, 1975, Muchinsky and Harris, 1977).

Table 1: Field experiments on gender discrimination: the effects of the gender ratio and status

Study and Method	Location and Time	Occupations	Measure	Finding
Gender ratio:				
Booth and Leigh (2010) (CS)	Brisbane, Melbourne, Sydney, Australia (2007)	Data-entrist (f, 85%), waiters (f, 80%), customer service employees (f, 68%) salesman (f, 69%)	Australian average (Australian Bureau of Statistics 2007)	Discrimination against men for data-entry and waitstaff, no discrimination for customer service and and sales
Riach and Rich (2006) (CS)	England (2003)	Chartered Accountants (n, 31%), computer analysts/ programmers (n, 21%), engineers (m, 5%), secretaries (f, 97%)	UK average (Office of National Statistics 2003)	Discrimination against women in (m) occupations and against men in (f) and (n) occupations, discrimination against men more pronounced
Weichselbaumer (2004) (CS)	Vienna, Austria (1998-1999)	Network technicians (m, 13%), computer programmers (m, 13%), accountants (f, 77%), secretaries (f, 97%)	Austrian average (Austrian Census 1991)	Discrimination against masculine and feminine women for network technicians and against men for secretaries, for others no discrimination between men, feminine women, and masculine women
Riach and Rich (1987) (CS)	Victoria, Australia (1983-1986)	Management Accountants (m, 9%), computer programmers (m, 23%), analyst programmers (m, 23%), computer operators (NI), industrial relations officers (NI), clerical workers (f, 68%), gardeners (m, 13%)	Victorian average (Australian Bureau of Statistics 1981)	Discrimination against women for computer programmers and gardeners, for others no discrimination
Levinson (1975) (AS) (telephone inquiries)	Atlanta, USA (1974)	Mostly secretaries/ receptionists (f), mostly security guards/ officers/ managers/ skilled workers (m)	NI	Discrimination against men in (f) occupations and against women in (m) occupations, discrimination against men more pronounced
Status:				
Neumark et al. (1996) (AS) (in-person inquiries)	Philadelphia, USA (1994)	Waiters (High, medium and low price/earnings)	Price level	Discrimination against women in high-price restaurants (job offfers and interview invitations) and against men in low-price restaurants (only w.r.t. job offers), no discrimination in medium-price restaurants
Firth (1982) (CS)	England (1978)	Articled clerks and qualified accountants working for professional accounting firms (noncareer jobs), unqualified personnel, qualified accountants working in industry and financial jobs (career jobs)	NI	Discrimination against women for qualified accountants working in industry and financial jobs, no discrimination for all other professions

Note: The table shows publications that compare discrimination across occupations with different shares of female employees as well as publications focusing on occupations with different social status. CS refers to a correspondence study while AS indicates an audit study. NI abbreviates "not indicated". (f) means that the authors classify the occupation as female dominated, (m) means that they classify an occupation as male dominated, and (n) means they classify an occupation as gender neutral, where the percentage refers to the ratio of female employees. Measure refers to the measure on which the authors base their categorization.

findings of Neumark et al. (1996) suggest that discrimination against women is higher for high-status occupations and lower, or potentially reversed, for low-status occupations (see also Table 1). The results of Firth (1982) can be interpreted as pointing in the same direction, although he finds discrimination against women only in two out of the three high-status professions. Equally, Riach and Rich (1987) note in their study that one of the two occupations for which they find discrimination is a high-status occupation, while the other is not. Unfortunately, not all of the studies report the measure of status on which they base their categorization. Given the few studies and the mixed results, the hypothesis that discrimination is more pronounced in high-status occupations calls for further scrutiny.

Larger firms have been found to discriminate less between men and women (Engels, 2015; Akar et al., 2014). One possible explanation is that evaluating larger groups of applicants leads to decisions that are less affected by group stereotypes (Bohnet et al., 2016). Another possible channel is that a more formalized and professional recruitment procedure might counteract gender discrimination. In addition, discrimination should be smaller if the total supply of applicants is small compared to the demand. Finally, parental background (the parents' education and profession) has been shown to influence the children's educational choices and achievements (Dustmann, 2004). Researchers have also found a high persistence of occupations between generations (Knoll et al., forthcoming; Jonsson et al. 2009). We can study whether the employers' evaluations contribute to these outcomes by studying the effect of the mother and father's profession on an applicant's evaluation.

Identifying discrimination

Just as field experiments, the vignette design allows us to observe the causal effect on the evaluation of being female or male. However, Heckman (1998) and Heckman and Siegelman (1993) have questioned the validity of field experiments that use correspondence methods.[7] Two points of their criticism remain valid for many recent field experiments, and we discuss that the factorial survey method can address one of them. Finally, we point out additional methodological benefits of our research design.

First, even if a researcher conducting a correspondence study is successful in making the applicants of the two groups under consideration equal with respect to the observed productivity characteristics, that is, the characteristics mentioned in the written application, a necessary assumption to identify discrimination is the equality of the average unobserved productivity-related factors of the two groups.[8] There should not be any differences in the mean of unobserved char-

[7] They have also criticized audit studies on the grounds that the testers acting as job applicants can influence the results. Often, the testers are not blind to the research question, thereby potentially causing experimenter demand effects. Note that correspondence tests relying on written applications do not suffer from this problem. They have also argued that any detection of discrimination in these studies only indicates that the average employer (and not the marginal employer) discriminates.

[8] Heckman and Siegelman make their argument for the case where unobserved productivity-related factors are factors that are unobservable to the researcher but visible to the employer and are taken into account in the

acteristics across the two groups. Otherwise, differences in outcomes might be due to differences in these unobservable productivity characteristics or due to group membership alone. Without assuming or ensuring that unobserved productivity factors are equal for both groups, researchers are not able to distinguish between the two.

The factorial survey method has in common with field experiments that we cannot differentiate between discrimination arising due to differences in the unobserved productivity characteristics or due to group membership alone. More generally, it is possible that the employers' estimates of the mean group productivity differ between men and women and that any differences in outcomes are due to these different expectations of employers upon seeing a male or female applicant. Thus, we can only identify discrimination that may be taste-based, statistical, or both.[9]

The second point raised by Heckman and Seligman concerns the variance of the unobserved productivity variable. Even if the mean unobserved productivity does not differ between the two groups under consideration, differences in the *variance* of the unobserved productivity between the groups can cause evidence of discrimination to be biased if the employers' hiring decisions are based on a cut-off rule (see Neumark (2012) for a detailed discussion). To see this, assume that both the average observed and unobserved productivity are equal across the two groups under consideration. Further assume that the variance of the unobserved productivity is higher for one group, say for women. If the researcher sets the observed productivity-related characteristics in the written application to a low value, and if the variance of unobserved characteristics is higher for women, women have a higher probability of being invited to an interview than men. In contrast, if the researcher designs applications with a high quality of observable productivity characteristics, the group with the lower variance in unobserved characteristics will receive more interview invitations. Thus, the observed discrimination can be an artefact of the study design.

Neumark (2012) proposes a statistical method to identify the level of discrimination despite this second problem. If a study has enough variation in the observable productivity characteristics, a heteroskedastic probit model can be used to infer the ratio of standard deviations of the unobserved productivity characteristics between the two groups and to back out discrimination. This procedure requires to assume that the coefficients on all observable applicant characteristics do not vary between the two groups under consideration. Neumark (2012) discusses examples for which this assumption is violated, e.g., if the productivity effects of schooling differ between two groups.[10]

Our estimate of discrimination between men and women is robust to Heckman and Seligman's unobserved variance critique for the following reasons. The estimate of discrimination is based on a 10-point evaluation scale and not on a cut-off value such as a binary invitation decision.

hiring decisions, while typically statistical discrimination refers to a situation where employers cannot observe an applicant's true productivity and hence partially rely on the average productivity of the group to which an applicant belongs (see below).

[9]For lab and lab-in-the-field experimental evidence regarding taste-based discrimination see Fershtman and Gneezy (2001) and List (2004).

[10]The advantage of the assumption is that it has testable implications.

As long as our 10-point scale fully covers all evaluations the employers want to make, which we deem reasonable, the employers' evaluations of the applicants are linear in productivity and our estimate of discrimination is robust to differences in the variance of unobserved productivity components between men and women.[11] However, even if this assumption was invalid, differences in the variance of unobserved productivity components between men and women are unlikely to cause our estimate to be biased for the following reason. As suggested in Neumark (2012), one way to address the unobserved-variance critique is to vary the level of applicant characteristics relevant to the hiring decision. If discrimination is found for different levels of the characteristics, the results cannot be due to differences in the variance of unobserved characteristics. The design of our vignette study allows us to simultaneously vary several applicant characteristics such as the duration of unemployment after leaving school or the average school grade. Significant coefficients show that these characteristics have an effect on hiring. Hence, our applicants differ with respect to their observable productivity level. In addition, employers from firms of eight different industries and 126 occupations evaluate our vignettes.[12] As the quality of an applicant depends on the requirements of the evaluator and on the occupation, we thereby consider a large variation in the quality of applicants.

Another advantage of the vignette design is that we can test whether Heckman and Siegelman's unobserved variance critique of field studies is warranted. Applicants are evaluated on a range from 1 to 10. If a difference in the variance of unobserved productivity components of men and women (resulting in a difference in the variance of productivity) is the reason for the discrimination against women found in earlier studies, this must be reflected in the distribution of evaluations. That is, if the variance of the unobserved productivity is higher for women than for men, we will observe a higher variance in the evaluations of women than of men.

Finally, the factorial survey design allows us to observe a finer measure of how applicants are evaluated. While it seems to be the case that firms evaluate applicants relative to a standard when they decide whether or not to invite a candidate to the next step of the recruitment process, the impact of an application on a candidate's employment probability at the stage at which field studies evaluate employer behavior is not necessarily binary. Although employers ultimately decide whether or not to hire a certain applicant, group membership might have an impact on the employer's evaluation beyond its impact on the call-back rate or the probability of being invited to an interview. An applicant from a minority might be invited to an interview though she or he might still be disadvantaged at later stages of the application process. By allowing employers to submit an evaluation on a scale from 1 to 10, we hope to capture more subtle differences between applicants.

[11] Note that our estimate of discrimination does not rely on the additional assumptions of Neumark's (2012) approach.

[12] In contrast, in the usual correspondence study applications are sent out for a small number of occupations (up to seven in the studies we are aware of).

Testing for statistical discrimination à la Aigner and Cain (1977)

We can use our dataset to shed some light on the nature of discrimination in our sample by testing for a certain form of statistical discrimination. In one of the classical models of statistical labor market discrimination by Aigner and Cain (1977), productivity is unobserved and employers have to form expectations about a worker's true productivity. The expected productivity of a worker is given by the weighted average of the mean productivity of the group to which the worker belongs and a measure of her individual productivity. In other words, since the true productivity is unobserved, employers rely partly on the group information. The weight on the individual productivity indicator is defined by the variance of the real productivity and the variance of the error when measuring the productivity indicator instead of the true productivity, and can be interpreted as the reliability of the productivity indicator. Statistical discrimination can arise due to differences in the average group productivities, but it can also arise due to differences in the reliability of the productivity indicator resulting in a worse signal strength (with average group productivities being equal). Intuitively, employers put more weight on the individual productivity indicator and less weight on the group mean for individuals belonging to the group for which the signal is more informative, which leads to an unequal treatment of equally productive applicants whenever the individual productivity indicator differs from the mean group productivity. For such differences in the reliability of the productivity indicator to arise, it can be assumed that the variance of the real productivity varies between groups, or that the variance of the real productivity is equivalent for the two groups but that the indicator is less precise in measuring real productivity.[13]

Two features of the design of our study enable us to test whether a difference in the signal strength induced by the productivity indicator being less precise in measuring real producitvity causes statistical discrimination. First, we vary the applicant characteristics quite substantially and hence expect to observe applicants that are considerably above the average productivity and applicants that are considerably below average. Second, under the additional assumption that the share of currently employed female apprentices is the decisive factor that causes the signal strength to vary, observing the share of currently employed female apprentices for all our employers means that we observe a proxy for the signal strength. For instance, firms with a high share of female employees have a better prior knowledge of the productivity distribution of women and, hence, the signal emerging from the productivity indicators is more precise. If this reasoning is valid, a better-than-average woman applying to these firms should be evaluated better than a man who is better than average and who has the same observable productivity characteristics. This is because employers put more weight on the woman's individual signal than on her group signal compared to a man. In contrast, a worse-than-average woman should be evaluated worse than a worse-than-average man who has the same observable productivity characteristics. Finally, the reverse

[13]The former assumption of different variances of real productivity between groups is comparable to Heckman and Siegelman's unobserved variance critique. Note again, however, that their criticism is based on an observed and an unobserved productivity component, which are both observed by employers, while in Aigner and Cain's model employers have to form an expectation.

reasoning applies for firms with a low share of female apprentices. We test whether the evaluations of men and women are consistent with this form of statistical discrimination.

3 Study design

The vignette study was embedded in an annual panel survey of firms engaged in apprenticeship training. In cooperation with the Company Panel on Qualification and Competence Development of the Federal Institute of Vocational Education and Training (BIBB), we included the vignettes as well as questions concerning the firms' recruitment procedures in the survey wave of 2014 (BIBB Training Panel 14). In total, 3,450 firms participate in the panel. Out of these, we randomly selected 680 firms to additionally take part in our factorial survey. About half of our firms have less than 100 employees. Some firms offer apprenticeships for several professions. We asked them to evaluate our fictitious applicants for their most frequent apprenticeship position. Among the firms in our survey, the most frequent apprenticeships are in technical professions (53%), followed by business or administrative professions (35%), and educational or nursing professions (12%).

The respondents of the computer-assisted personal interviews (CAPI) were firm owners, managers, and employees involved in human resources. They were first given a number of questions regarding the hiring process. Then they were asked to evaluate short descriptions of applicants (vignettes) for an apprenticeship in the occupation for which their organization trains most people. Within the vignettes, we varied the applicants' attributes (dimensions). The method, also known as a factorial survey experiment, embeds randomized treatments in a survey. It allows for studying the relative importance of a number of applicant characteristics simultaneously. Thus, we can determine which information about the applicant is actually being used. Moreover, it is possible to measure the influence of a single variable and of combinations of variables on the evaluations. All respondents had to evaluate five vignettes, each of them describing a fictitious applicant, with respect to whether he or she would reach the next step of the hiring procedure. Answers had to be provided on a scale from 1 (very unlikely) to 10 (very likely). Overall, we attempted to collect 3,400 evaluations of the fictitious applicants.

In comparison to correspondence tests, the factorial survey method relies on hypothetical, not real decisions. Hence, the answers might differ from actual behavior. Similarly, the answers might be affected by demand effects when respondents try to please the interviewer or try to provide socially acceptable answers.[14] In order to limit such concerns with respect to gender as our main variable of interest, the five fictitious applicants described in the vignettes that every respondent received from us were of the same gender. Demand effects are also minimized by the fact that respondents were not asked directly, for example, for the role of gender or grades for their

[14] This might not only bias our baseline findings, but also the findings of our correlational analysis. For instance, the desire to give socially acceptable answers might be differently strong for occupations with a different share of female apprentices.

Figure 1: Example of a vignette as shown to the respondents (translated from German)

Gender: Female	**Education:** Intermediate school-leaving degree 2012
Date of birth: Nov. 3, 1995	**Final grade point average:** 3.4
Father's occupation: Employee in firm	**Social behavior according to school reports:** Mostly good
Mother's occupation: Elderly care nurse	**Unexcused absent days:** Three
Current activity: Temporary job	**Activity since finishing school:** One-year pre-vocational training and apprenticeship (discontinued); since Dec. 2013: temporary job

Note: The respondents were asked: *"How likely is it that this applicant is invited to the next step of the recruitment process?"* Possible answers were between 1 (very unlikely) to 10 (very likely).

evaluations, but only indirectly by being asked to evaluate fictitious applicants.

Figure 1 displays an example of a vignette as it was shown to the respondents (translated from German).[15] Applicants have completed or planned to complete the intermediate schooling degree. All other dimensions were varied. The year of birth was either 1993, 1995 or 1997 such that the applicants were aged 16, 18 or 20 at the start of the apprenticeship. For those born in 1993 and 1995, we varied the information about their activities since leaving school. Either they participated in a one-year pre-vocational training, followed by an apprenticeship that was discontinued, or they did not provide any information on their activities since leaving school two or four years ago.[16] For the applicants born in 1993 and 1995, we further indicated that they were currently working in a temporary job. For the applicants born in 1997, we indicated that this was not applicable since they were still in school.

The average grades from school were either 2.8 or 3.4. Both are satisfactory and typical of school leavers with an intermediate degree, but a grade point average (GPA) of 2.8 is clearly better than 3.4.[17] Applicants get grades for their social behavior in school, and we varied this between very good and medium. The transcript from school also includes absent days without excuse, and we varied this between none and three days.

[15] For a vignette in German and the introductory text shown to the respondents see Appendix A.1.

[16] Pre-vocational training measures are government-sponsored programs aimed at preparing students for apprenticeships. The programs are non-selective. Baethge et al. (2007) and Fitzenberger et al. (2015) provide a description of the various measures of the transition system and their economic significance.

[17] Students are not allowed to finish more than two courses with a grade of 5, and have to repeat the school year otherwise. Given this constraint, an average of 3.4 is a relatively poor GPA. On the other hand, students with a GPA higher than 2.8 often decide to get a high school degree.

Finally, in order to investigate whether the socio-economic background of the applicant matters, we chose different professions of the father. He was either a warehouse clerk, an insurance salesman, a teacher, or an employee in the respective firm. Many firms have rules according to which children of employees automatically pass the first selection step, and we therefore expect these applicants to pass the hurdle with a higher probability. The other three professions differ with respect to salaries and educational requirements.[18] The mother had one of two professions of similar status, salary, and required education, namely nursery-school teacher and elderly-care nurse.[19]

4 Results

Our main dependent variable is the evaluation of the fictitious applicants on a 10-point scale. The response rate of the 680 firms selected for our survey is more than 98% with only 47 out of 3,400 evaluations of fictitious applicants missing. We exclude from the analysis respondents who indicate that they do not know the recruitment procedure of their firm, respondents who state that their firm does not have a recruitment procedure, and respondents who report that their firm does not review and examine the application documents. Overall, this leaves us with a sample of 3,164 evaluations made by 636 respondents.

The BIBB provides us with sampling weights for our set of firms to correct for imperfections of the sample with respect to the full sample of firms training apprentices in Germany. We conducted all estimations both with and without the sample weights. Most results are qualitatively unchanged. We state all results based on the unweighted estimations. However, we also provide the results using the sampling weights whenever they deviate from the unweighted results with respect to the significance of an effect.

4.1 Is there a gender difference in evaluations?

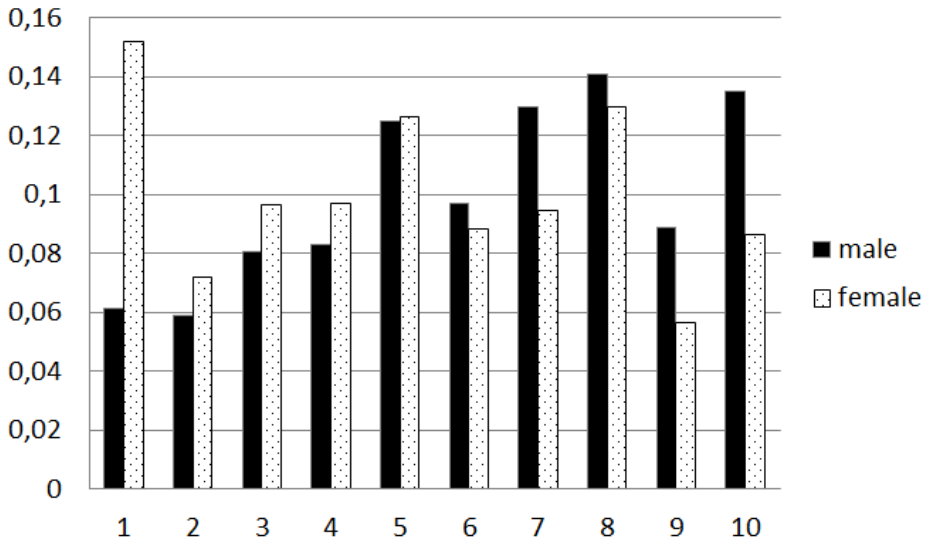
For a first impression of the evaluations provided by the respondents, Figure 2 shows the distribution of evaluations for male and female applicants. Apart from indicating that the respondents use the full scale of possible evaluations, the figure also reveals a difference in the evaluation of men and women, with more frequent positive evaluations of men compared to women. Women receive bad evaluations (1-5) more often than men, while men receive good evaluations (6-10) more often than women. The bars for the worst (1) and the best evaluation (10) display the largest differences between women and men. For instance, 241 female applicants receive an evaluation of 1, indicating that an invitation to the next recruitment step is very unlikely, while only 97 male applicants

[18] In Germany a warehouse clerk earned, on average, between 12.98 and 15.55 Euro per hour in 2014, while a teacher earned 24.78, and an insurance salesman earned between 23.74 and 28.98 Euro (Federal Statistics Office Destatis, 2016). Working as a teacher requires a university degree in Germany while insurance brokers and (to a lesser degree) warehouse clerks usually complete an apprenticeship.

[19] Nursery-school teachers earned, on average, between 13.60 and 16.43 Euro in Germany in 2014 while elderly-care nurses earned 14.42 Euro (Federal Statistics Office Destatis, 2016).

receive this evaluation. Only 137 female applicants receive an evaluation of 10, indicating that the invitation is very likely, while 213 men receive this evaluation. Male applicants receive an average evaluation of 6.14 while female applicants receive an average evaluation of 5.21 such that the male applicants' evaluations are on average 0.93 points better than those of female applicants. This difference in means is significant (t-test, $p<0.001$).

Figure 2: Evaluation of Fictitious Applicants by Gender



Note: Frequency of each evaluation for male and female applicants.

Our vignettes are chosen such that the applicant characteristics are equal for men and women in expectation. Thus, if the randomization worked, the observed mean discrimination against women should be unchanged if we regress the evaluations on a gender dummy and all other vignette dimensions (applicant's age, mother and father's occupation, average grade, social behavior, number of absent days unexcused, and the gap after school). Since one respondent evaluated five vignettes, we use a random-effects (RE) estimation and take into account the dependencies of the answers at the level of the respondents (firms).[20] We find that women are evaluated 0.90 points worse than men, which is highly statistically significant (see Table 2). It is also economically relevant: Supposing a linear effect of an applicant's GPA on the evaluation, being a man instead of a woman makes up for almost one grade level from school. The only vignette dimension that has a stronger impact on the respondents' evaluations is a four-year gap after school. Since applications from older applicants indicate that they had issues at their career entry, it is not surprising that this variable is relevant for the employers' evaluations. Besides, Table 2 indicates that several other applicant characteristics matter for the employers' evaluations, and the coefficients have the expected signs.

[20]For the estimation using the sample weights, we use a multilevel mixed-effects model. Our results using the unweighted sample are robust to using a multilevel mixed-effects model.

Table 2: Baseline results

	Coefficient	SE
Female	-0.90**	0.174
Age 18	-0.89**	0.085
Age 20	-0.93**	0.089
No info about gap	-0.05	0.079
Worse GPA	-0.77**	0.063
Intermediate social behavior	-0.45**	0.063
Three absent days	-0.81**	0.063
Mother elderly-care nurse	0.11	0.063
Father's occupation		
Insurance salesman	-0.02	0.084
Teacher	-0.03	0.084
Employee in firm	0.30**	0.105
Constant	8.49**	0.177
Observations	3164	

Notes: ** $p < 0.01$, * $p < 0.05$. Coefficients and standard errors are obtained from a RE estimation from the evaluation on the listed variables. Omitted categories are age 16, information about gap, warehouse clerk, nursery-school teacher, better GPA, good social behavior, and no absent days.

Do these results regarding discrimination hold up to Heckman and Seligman's unobserved variance critique? A difference in the variance of unobserved productivity characteristics between men and women that results in a difference in the variance of productivity as assessed by the employers should be reflected in a difference in the variance of evaluations. We find that the variance of the evaluations is statistically significantly larger for women than for men.[21] Such a difference can cause the results of field experiments to be biased unless the study relies on the estimation procedure proposed by Neumark (2012). Our own estimate of discrimination between men and women is robust to the finding of differences in the variance due to the fine-grained evaluations of the applicants that we collected. Note that an unbiased estimate also requires that the 10-point scale fully captures the employers' valuations. If respondents wanted to evaluate some women worse and some men better than possible with the scale, our mean estimate would underestimate the true penalty in evaluations for the women. Note that there are many evaluations of 1 for women and of 10 for men. However, it is unlikely that we set a particularly low or high

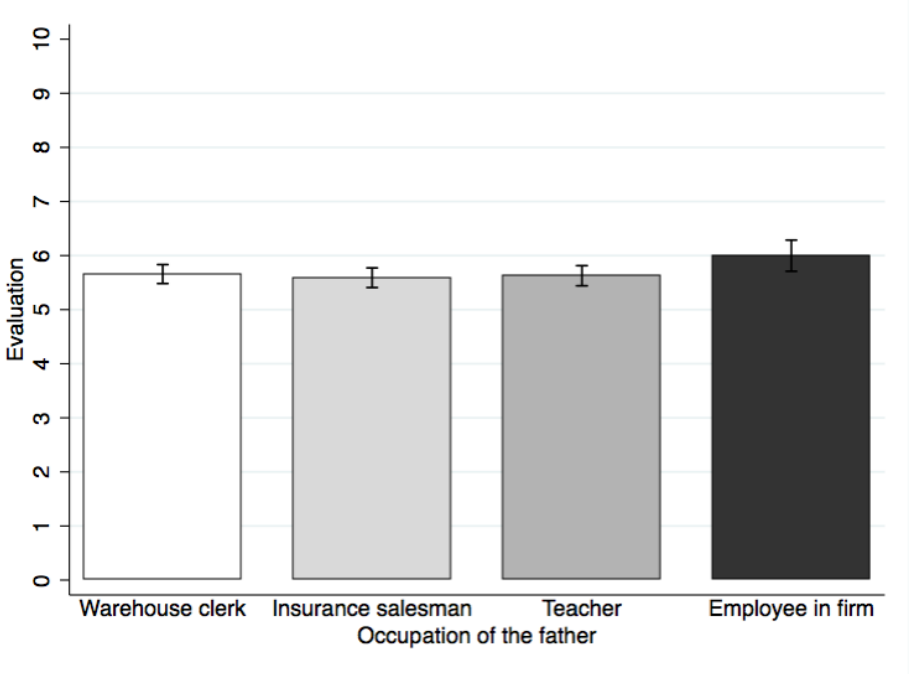
[21] An F-test of equality of variances strongly rejects the null of equality of variances. We perform Levene's test and the Brown-Forsythe test that are both robust to nonnormality of the data, and we also find that variances are different.

standard when designing the vignettes, since gender seems to be the main factor leading to extreme valuations: Weaker male applicants are only seldom evaluated poorly (only about 5 percent of men receive an evaluation of 1 or 2), while even good female applicants are not able to score high (only between 5 and 8 percent of women receive an evaluation of 9 or 10, respectively).

4.2 Parental background

It is common for applicants for apprenticeships in Germany to indicate the occupation of their parents on the CV. We vary the mother and father's occupation in order to analyze whether there is discrimination with respect to the applicant's socio-economic background. We find no statistically significant difference in the mean evaluations comparing applicants whose fathers are warehouse clerks and insurance salesmen (t-test, p=0.597), comparing applicants whose fathers are warehouse clerks and teachers (t-test, p=0.814), and comparing applicants whose fathers are teachers and insurance salesmen (t-test, p=0.778), see Figure 3.

Figure 3: Evaluation of fictitious applicants by occupation of father



Note: Average evaluations (along with 95% confidence intervals) of applicants for each occupation of the father.

There are two possible reasons for this finding. First, it could be that discrimination based on socioeconomic background is not an issue in the market for apprenticeships. Alternatively it might be that, although employers discriminate based on social background, they do not evaluate applicants differently based on their parental background in our vignette study, because they think it is socially unacceptable to do so. Note that while the respondents in the sample only evaluate either male or female applicants, we vary the parental background between the five vignettes of

one respondent. Hence, respondents might be reluctant to differentiate between applicants based on the socio-economic background of the parents.

Applicants whose father is employed in the firm receive an evaluation that is 0.37 points better than applicants whose father works in one of the three other occupations, which is significant (t-test, p=0.011).[22] However, using the sample weights this difference becomes insignificant. For the mother's occupations of nursery-school teacher and elderly-care nurse, just as for the occupation of the father, we do not find a significant difference in the mean evaluations between applicants (t-test, p=0.153).

4.3 Applicant quality

Gender discrimination may depend on the perceived ability of an applicant. We study this possibility in two different ways.

First, we compare gender discrimination between applicants with good (2.8) and medium (3.4) grades. Figure 4 plots the mean applicant evaluations separately by gender and GPA. We see that the applicants with the better GPA are evaluated better than the applicants with the worse GPA. There is a gender difference in evaluations for both levels of the GPA, and the gender differences do not seem to vary between grade levels. Using a t-test, we find that the gender difference is highly statistically significant for applicants with the better GPA (t-test, p<0.001) as well as for applicants with the poorer GPA (t-test, p<0.001) and comparable in size (0.95 and 0.89; t-test, p=0.736). Using the sample weights, the gender difference in evaluations for applicants with the better GPA is 0.62 and fails to be significant at the 5 percent level, but the difference remains significant for applicants with the worse GPA.

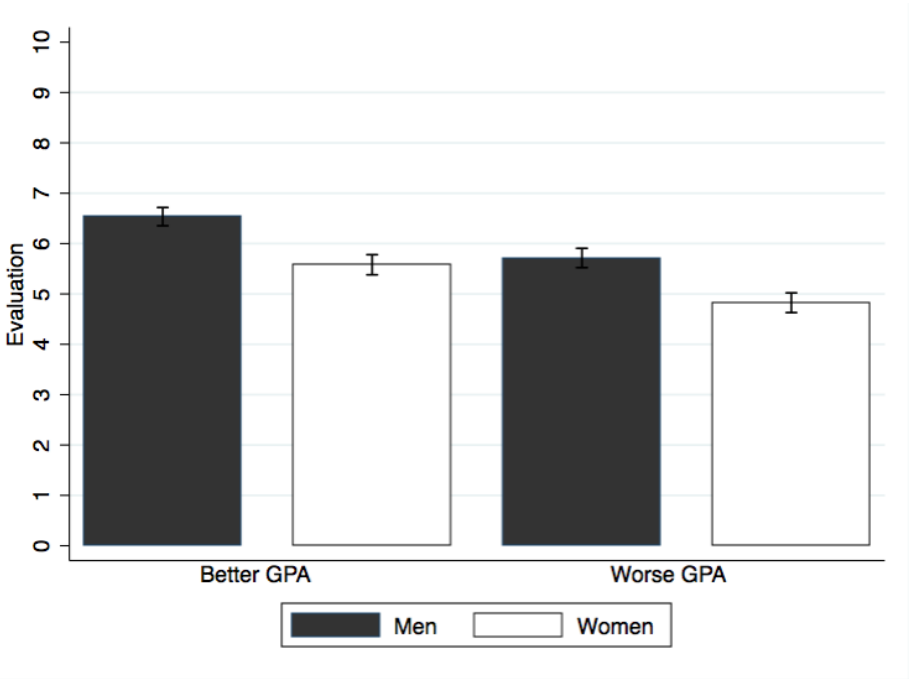
Second, we can compare the gender difference between applicants grouped according to how they are evaluated by the respondents. Figure 2 suggests that the gender effect is more pronounced for applicants who are evaluated as very likely to be invited to the next step and for those evaluated as very unlikely, and that the gender difference is smaller for intermediate applications. To test this, we define good [bad] applicants as those who are evaluated as good [bad] and use a random-effects ordered probit estimation to estimate the marginal effect of being female on the probability of observing each of the 10 outcome categories, employing the same controls as before (the vignette dimensions).[23] Being a woman significantly increases the probability of receiving an evaluation of 1 (the lowest possible evaluation) by 6.5 percentage points. This gender difference declines over the outcome categories such that the effect of being a female applicant is small in absolute terms but significant for the intermediate evaluations of 5 and 6, amounting to an increase of 0.8 percentage points for being in category 5 and amounting to a 0.2 percentage point decrease for category 6.

[22] If each one of the three occupations is tested separately against "father is employee of the firm," we also find that the latter leads to a significantly better evaluation.

[23] See Appendix A.2, Table 5. Note that we refrain from using RE ordered probit regressions in the rest of the paper for ease of interpretation.

The gender difference is largest for an evaluation of 10 (the highest possible evaluation). Being a woman significantly decreases the probability of this evaluation by 6.7 percentage points. Hence, while discrimination is not restricted to only good or only bad applicants, it is highest at the extremes of the distribution of evaluations. When we use the sample weights, the marginal effect of gender fails to be statistically significant at the 5 percent level for several outcomes (1, 5, 6, 9 and 10), although the coefficients remain constant in size. The findings suggest that gender differences in evaluations are not restricted to certain ranges of the ability distribution albeit the size and significance of discrimination vary.

Figure 4: Evaluation of fictitious applicants by gender and grade



Note: Average evaluations of male and female applicants with a better GPA (2.8) and a worse GPA (3.4) along with 95%-confidence intervals for the means.

We can also test whether the observed gender difference in evaluations is due to statistical discrimination caused by differences in the precision of the observable productivity components (Aigner and Cain, 1977). Suppose that a difference in the signal quality of the individual productivity components between men and women is induced by the fact that observable productivity components are less precise in measuring real productivity for one gender and not by the variance of the real productivity beeing different between men and women.[24] If this difference in the signal quality causes the gender difference in evaluations and if the share of currently employed female apprentices is the decisive factor for differences in signal strength, a better-than-average

[24]Note that our results indicated that the variance of real productivity might be higher for women than for men (see Section 4.1). This, however, only holds true in the set-up of Heckman (1998) and Heckman and Siegelman (1993) according to which employers evaluate applicants based on their observed and unobserved productivity characteristics.

[worse-than-average] woman applying to firms with a high share of female apprentices should be evaluated better [worse] than a better-than-average [worse-than-average] man with the same observable productivity characteristics applying to the same firms. The opposite holds true for firms with a low share of currently employed female apprentices. To test this, we match information on the share of women currently in an apprenticeship occupation to our data.[25] We first consider occupations with a share of 67 to 100 percent female apprentices, that is, occupations for which we assume that the signal for the individual component is of higher quality for women than for men. We define a better-than-average applicant as an applicant with the more desirable characteristics of being 16 years of age, having the better GPA, and zero absent days.[26] For this selected sample, we conduct a t-test on the difference in mean evaluations between men and women. We find the difference to be positive (indicating that men are evaluated better) but not statistically significantly so. In turn, we perform a t-test for applicants who we expect to be below average in quality, defined as applicants with less desirable characteristics. Again, we find the difference between men and women to be positive but not significantly different from zero. Repeating this task for male-dominated occupations, we find statistically significant and positive differences in evaluations between men and women not only for better-than-average applicants – as the theory would suggest – but also for worse-than-average applicants – contradicting the theory.[27]

Summing up, we find statistically significant gender difference in evaluations between men and women for both levels of the GPA. Moreover, discrimination is not restricted to certain parts of the distribution of evaluations. The gender differences in evaluations are not driven by statistical discrimination based on gender differences in the precision of signals of the individual productivities. Thus, the observed gender discrimination can be due to statistical discrimination caused by differences in the average group productivity, by differences in the variance of the true individual productivity, or by taste-based discrimination.

4.4 Characteristics of occupations and firms

Besides assessing the influence of the vignette characteristics on gender discrimination, our dataset also allows us to investigate the potential firm-related or occupation-related mechanisms behind the observed differential evaluation of men and women. In the remainder of the paper, we will investigate these potential moderators by exploiting our rich dataset. We sequentially interact each variable we consider with the gender dummy and control for the dimensions of the vignettes in a random-effects (RE) estimation. In Appendix A.2, Table 6 to Table 13, we report on the detailed results of all estimations.

[25] We use the German classification of occupations (KldB 2010). The data on the share of female apprentices by profession is provided by the Federal Institute for Vocational Education and Training (BIBB) and the Federal Statistics Office Destatis. See Appendix A.3, Table 14, for an overview of all external data sources used.

[26] These three vignette characteristics are the most important according to our baseline results in part 4.1.

[27] Using the sample weights, the former is not statistically significant, which is also in contrast to the theory.

4.4.1 Different industries and occupations

We first analyze how the observed difference in evaluations between men and women differs between occupation types and industries. In the survey, respondents were asked to classify their most common apprenticeship occupation as "business/administrative", "technical", or "educational/nursing" apprenticeship occupations. We regress the evaluation on a gender dummy, dummies for occupations classified as "business/administrative" and "educational/nursing" and the interactions between these two occupation classifications and the gender dummy in a RE regression. We additionally control for all vignette characteristics.

We find that male applicants are evaluated better than female applicants for technical occupations and for educational or nursing occupations. For these occupations being a woman decreases evaluations by 1.23 and 1.09 points, respectively. Although for business or administrative occupations the effect of being a women is also negative, amounting to 0.42 points, it is not significantly different from zero.

As an alternative to classifying the apprenticeship occupations, we can also group the firms into industries. Our dataset contains firms from eight industries, and we observe, on average, 396 evaluations of applicants by 79 respondents per industry.[28] Figure 5 shows the difference in evaluations of male and female applicants by industries along with 95% confidence intervals obtained by regressing the evaluation on a gender dummy, industry dummies and the interactions between the industry dummies and the gender dummy (controlling for all other vignette characteristics) using a RE regression.[29] We find a gender difference in evaluations that is significantly different from zero for "Manufacturing", "Construction" and for "Public Service & Education," amounting to 0.89, 2.40 and 1.14, respectively. For the other industries, the coefficient of gender is not significantly different from zero, although women are evaluated worse than men in all industries.

Note that the manufacturing and construction industries are among those with the lowest shares of female apprentices with 23 and 7 percent, respectively. In the next section, we consider the role of this share for gender discrimination.

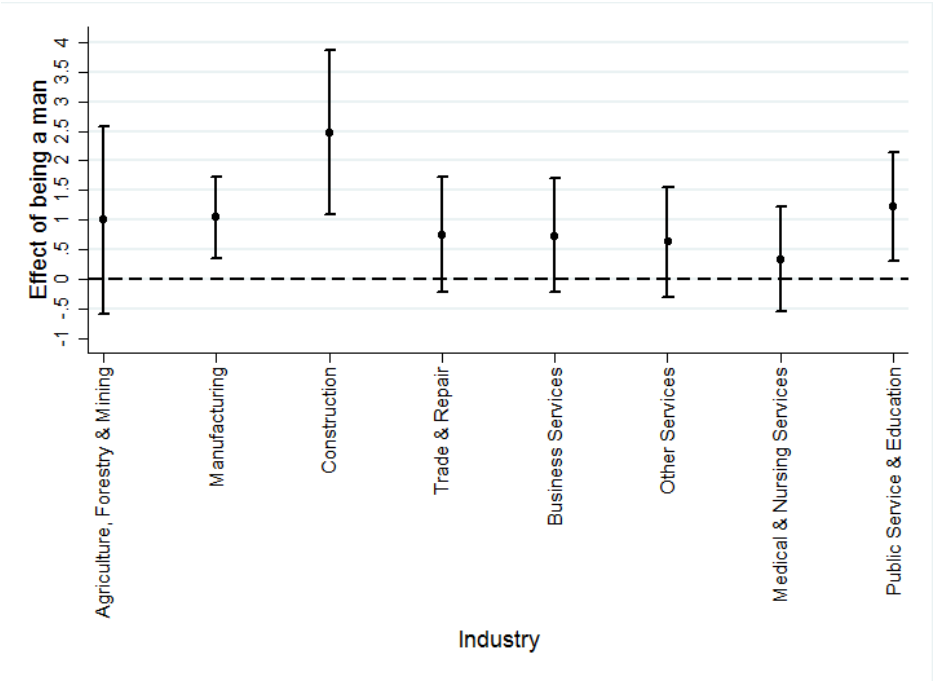
4.4.2 Male- versus female-dominated professions

One potential moderator of gender discrimination is the share of women and men in an occupation. Employers may prefer hiring the dominant gender because they expect this to be conducive to the work atmosphere. To study how gender discrimination depends on the share of employees in the occupation that are female, we information about the share of women currently in an apprenticeship occupation to our data. In Appendix A.4 we also report on regressions using an

[28] The classification of firms into industries is based on the German Classification of Economic Activities, 2008 (WZ 08).

[29] Our results are almost exactly the same with and without controlling for all other vignette characteristics. This indicates that even within the eight categories our randomization was successful and applicant characteristics, on average, equalize across gender.

Figure 5: Gender difference in evaluations by industry

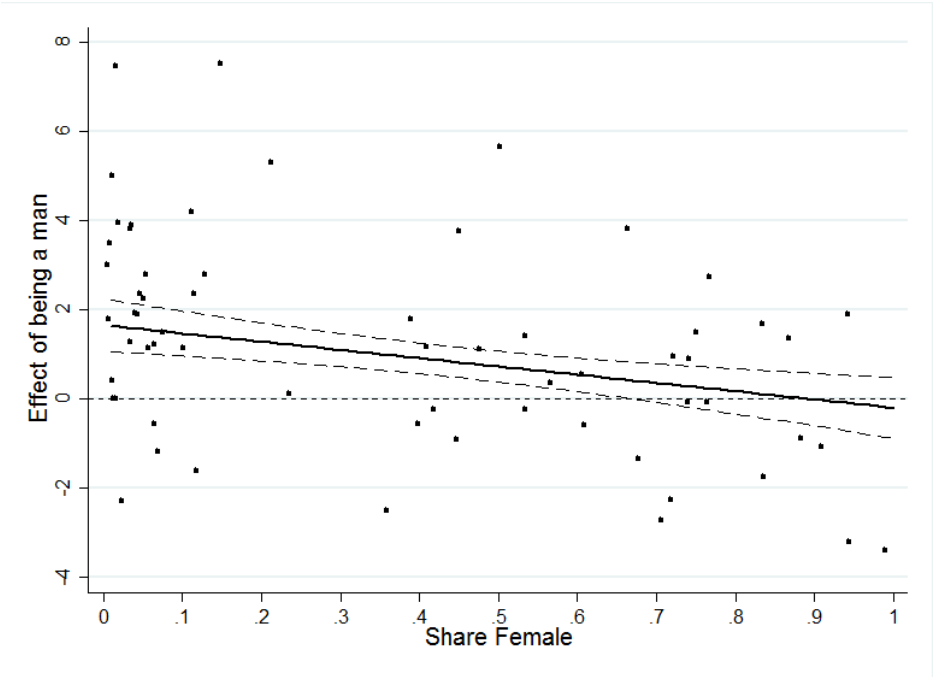


Note: The figure presents the effect of being a man along with 95% confidence intervals obtained by a RE regression of the evaluation on a gender dummy, industry dummies, the interactions between the industry dummies and the gender dummy and all other vignette characteristics.

alternative measure, namely the share of all female apprentices in the apprenticeship occupation in a firm that are self-reported by the respondents. This analysis leads to qualitatively similar results.

Figure 6 plots the difference in mean evaluations between male and female applicants against the proportion of female employees in the respective occupation in Germany. The dots suggest that the smaller the proportion of female apprentices, the worse the relative evaluation of women compared to men. The figure also shows the results from a RE regression of the evaluation on a gender dummy, the share of women currently employed in an apprenticeship occupation, the interaction between the share of women and the gender dummy, as well as the other vignette characteristics. There is a negative relationship between the share of current female apprentices and the evaluation of men relative to women. Increasing the share of female apprentices by 50 percentage points is associated with a decrease in the absolute amount of discrimination, namely a decrease in the effect of being a man by 0.93. This effect is highly statistically significant. Moreover, the difference in evaluations between male and female applicants is significantly different from zero at the five percent level for jobs with a share of women below 68 percent, but not significantly different from zero for jobs with a share of women above 68 percent. For example, we find no evidence of discrimination against women for the occupation of nursing where the share of women is 77 percent.

Figure 6: Gender difference in evaluations by share of women



Note: The dots represent gender differences in the evaluation of applicants for apprenticeship occupations with a given share of women. The solid line shows the effect of being a man obtained by a RE regression of the evaluation on a gender dummy, the share of women currently employed in an apprenticeship occupation, the interaction between the share of women and the gender dummy and all other vignette characteristics. The dashed lines indicate 95% confidence intervals for this effect.

Following the literature, we also split the occupations into three groups. We call those occupations with a national average of 0 to 33 percent female apprentices male-dominated, those with a share of 33 to 66 percent gender balanced and those with a share of 67 to 100 percent female apprentices female-dominated.[30] We regress our evaluations on a gender dummy and both the interaction terms of male-dominated jobs and gender as well as female-dominated jobs and gender, that is, we interact both the indicator for female-dominated and for male-dominated jobs with the gender dummy within one regression. For gender-balanced apprenticeship occupations, we neither find significant evidence of discrimination against men nor against women (see Table 3). There is no significant difference in discrimination between female-dominated and gender-neutral occupations. Hence, we do not find evidence of discrimination against men in female-dominated jobs in contrast to results in the literature. For male-dominated jobs, the effect of being a woman is statistically significantly different from zero and amounts to -1.51 points. This is evidence of discrimination against women in male-dominated jobs.

[30]The results are robust to using cut-offs at 30 percent and 70 percent as in Weichselbaumer, 2004.

Table 3: Discrimination and gender ratio of occupations

Variable	Coefficient	SE
Female	-0.45	0.344
Female+Female*Female-dominated	-0.37	
Female+Female*Male-dominated	-1.51**	

Notes: The table displays the coefficients and the standard errors from a random effects regression of the evaluations on a gender dummy, an indicator for female-dominated occupations, an indicator for male-dominated occupations, interaction terms between the gender dummy and the two occupation class indicators ("Female*Female-dominated" and "Female*Male-dominated"), and other vignette characteristics.

4.4.3 High- versus low-status occupations

The literature presents evidence that discrimination against women is more severe for occupations associated with a high social status (for an overview, see Riach and Rich, 2002). We analyze how gender discrimination varies with several variables related to occupational status. In particular, we investigate the relationship between gender discrimination and the average salary, the typical school-leaving qualification required for an apprenticeship, and two indices of occupational status.

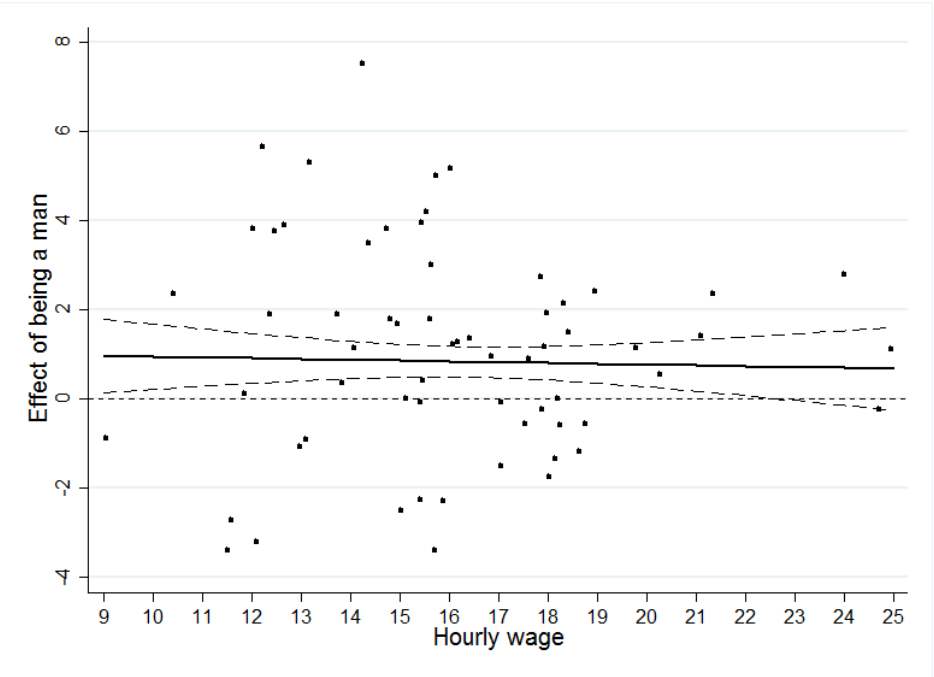
Salaries

For the analysis of the salaries, we merge average hourly wages to our professions via the KldB 2010. The wages are national averages provided by the Federal Statistical Office (Destatis). We end up with 117 different hourly wages, ranging from 9.05 Euro per hour to 24.95 Euro per hour. Our dataset also contains information on the average monthly wages of all workers in a firm. Based on these data, we construct two additional measures of salaries. The analyses using these data can be found in Appendix A.5. The results are in line with the main results using hourly wage data for the different occupations presented in this section.

Figure 7 displays how the difference in evaluations between men and women varies with the hourly wage. The graphical evidence suggests no clear linear relationship. The figure also contains the gender effect for different wages obtained by regressing the evaluation on a gender dummy, the average hourly wage of an occupation, their interactions, and the vignette characteristics. We see that the negative effect of being a woman slightly decreases in absolute terms with higher wages, and the effect of being a man is not significantly different from zero for very high wages. Overall, there is no clear linear relationship between the amount of discrimination and the salary of a profession.

To allow for a non-linear effect, we divide the wage distribution at the 0.33-quantile and the 0.66-

Figure 7: Gender difference in evaluations by hourly wage of occupation



Note: The dots represent gender differences in the evaluation of applicants for a given wage level. The solid line shows the effect of being a man obtained by a RE regression of the evaluation on a gender dummy, the hourly wage of an occupation, the interaction between the hourly wage of an occupation and the gender dummy and all other vignette characteristics. The dashed lines indicate 95% confidence intervals for this effect.

quantile and interact the gender dummy with indicators for each value range in a RE regression. There is strong evidence of gender discrimination for firms that pay low wages, i.e., for firms paying a wage between 9.05 and 15.41 Euro per hour. For these firms, being a woman instead of a man lowers the evaluation by 0.74 points, which is significantly different from zero. For occupations paying intermediate wages of between 15.43 and 17.89 Euro, we also observe significant discrimination against women, amounting to 1.19 evaluation points. Finally, firms paying high wages evaluate women 0.62 points worse than men, which again is significant. Hence, we do not find strong evidence that discrimination systematically varies with the average salary of a profession.

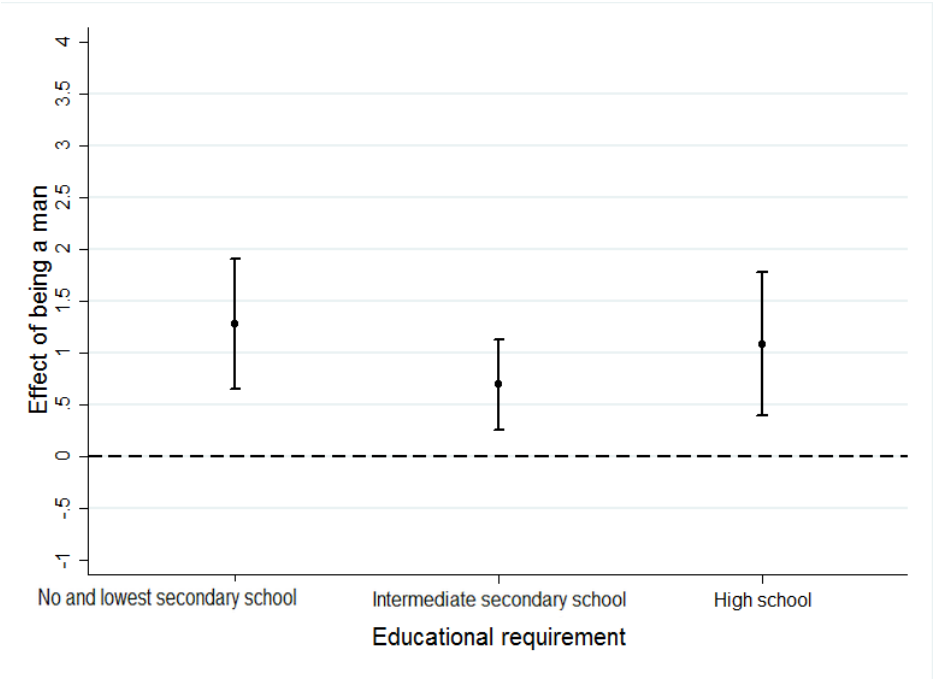
Education requirements

We now consider how discrimination varies with the educational requirements of a profession. After primary school, German students attend different secondary school tracks. These tracks differ in their lengths and educational orientation. We can differentiate between evaluations made by firms for which the typical school-leaving qualification of a currently employed apprentice is (i) no school-leaving qualification or a school-leaving qualification below or equal to lower secondary school

(Hauptschule), (ii) intermediate secondary school (Realschule), or (iii) high school (Gymnasium).[31]

Figure 8 shows the results from our correlation analysis. Interacting gender with our indicators of the school-leaving qualification in a RE regression, we find that the difference in evaluations between men and women is significant for all three categories. In particular male applicants are evaluated 1.28 points better when a lower secondary degree is required, 0.69 points better by firms requiring an intermediate school degree, and 1.08 points better when a high school degree is necessary. When using the sample weights for the data, the only difference that emerges is that the gender difference for the lowest school track becomes very small and is no longer significant.

Figure 8: Gender difference in evaluations by educational requirement



Note: The figure presents the effect of being a man along with 95% confidence intervals obtained by a RE regression of the evaluation on a gender dummy, the school-leaving qualification required, the interactions between the required school-leaving qualification and the gender dummy and all other vignette characteristics.

Our dataset contains another proxy of the typical education required for a job, based on a self-reported measure of how many of the firms' apprentices hold each of the various school-leaving qualifications. We report on this analysis in Appendix A.6. Based on both measures, we find that there is discrimination against women for professions and firms with low, medium, and high educational requirements, with no clear differences in the amount of discrimination between these occupations.

[31] In the case that a respondent indicates that two school-leaving qualifications are equally likely, we randomly break ties. If a respondent indicates that more than two school-leaving qualifications are equally likely, the observation is discarded. Since only four respondents indicate that their apprentices typically have no school-leaving qualification, we do not distinguish between no school-leaving qualification and the lowest school-leaving qualification ("Hauptschule").

Occupational Status

Various measures of occupational status have been discussed in the literature (Ganzeboom and Treiman, 2003). We employ the International Socio-Economic Index of occupational status (ISEI-08) which is a socioeconomic measure of occupational status.[32] It is based on the assumption that occupational status determines how education is transformed into earnings.[33] The index was obtained using data from 2002 to 2007 on 200,000 individuals from more than 40 countries (Ganzeboom, 2010). In our sample the ISEI is lowest for unskilled agricultural laborers who have a score of 11.74 and highest for application programmers with a score of 74.66. The average score is 36.92 and applies, for instance, to electrical mechanics.

Figure 9 displays the average differences in evaluations between male and female applicants for different values of the index of socioeconomic status. Male applicants are evaluated better than female applicants for the majority of status scores. The figure further suggests a negative relationship between the difference in evaluations of male and female applicants and the socioeconomic status of an occupation. This is in line with the result we obtain when we interact the socioeconomic measure of occupational status with gender and run a RE estimation on a gender dummy, the ISEI, the interaction between them as well as the vignette characteristics (see also Figure 9). There is a sizable negative baseline effect of being female on the evaluations, while an increase in the status of an occupation is associated with a decrease in gender discrimination as indicated by a significant coefficient on the interaction term. In line with this, we find a statistically significant discrimination of 1.24 points for waitresses but no evidence of discrimination for application programmers (effect size=0.10). Using the weighting factor, the coefficient on the interaction term does not change in size, but fails to be statistically significant at the 5 percent level.

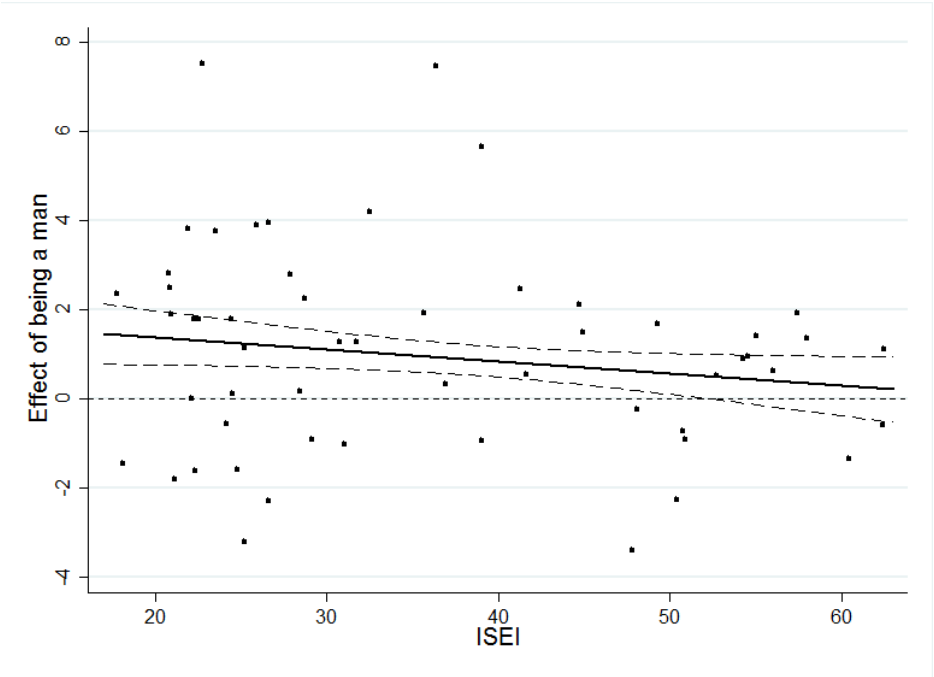
We employ a second indicator of occupational status, the Standard International Occupation Prestige Score (SIOPS) which is a measure of occupational prestige (Treiman, 1977). It is obtained by asking individuals to assess the popularity of occupations.[34] In our dataset, the scores of occupational prestige range from 16, assigned to fast food sellers, to 60 for denturists. Similarly, unskilled agricultural laborers and waiters score very low when it comes to occupational prestige

[32]The ISEI-08 is the International Socio-Economic Index of occupational status estimated for the ISCO-08, a classification of occupations based on skill level and skill specialization. We rely on the conversion provided by Ganzeboom and Treiman (2017). Since in our dataset firms are categorized according to the KldB 2010, we first add the ISCO-08 to our data, that is, we assign one unit group of the ISCO-08 (four-digit) to every occupational type of the KldB 2010 (five-digit). We do this by using the conversion key provided by the Federal Employment Agency, see https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Arbeitshilfen/Umsteigeschluessel/Umsteigeschluessel-Nav.html.

[33]The ISEI of an occupation is found by scaling occupational groups as a mediating variable such that the direct effects of education on earnings are minimized and effects of education on earnings through occupation are maximized. This leads to two coefficients, one linking occupational status to education and one relating occupational status to income that are used as weights to produce status scores based on the average level of education and earnings of an occupation.

[34]Individuals are asked to rate, for instance, the social prestige, the social standing, or the desirability of an occupation. Treiman matched these prestige scores to occupations by using data from 60 countries. This measure of occupational prestige might capture additional dimensions of occupational status that are not captured by the socioeconomic measure.

Figure 9: Gender difference in evaluations by status of occupation



Note: The dots represent gender differences in the evaluation of applicants. The solid line shows the effect of being a man obtained by a RE regression of the evaluation on a gender dummy, the ISEI, the interaction between the ISEI and the gender dummy and all other vignette characteristics. The dashed lines indicate 95% confidence intervals for this effect.

whereas drafters and opticians score very high.

If we interact this measure of the prestige of an occupation with gender in our RE estimation, raising the SIOPS leads to a decrease in discrimination, which is in line with the results for the ISEI. However, the coefficient of the interaction term is not significant. Considering as examples the scores of waiters or waitresses and opticians yields the following findings. There is significant discrimination against women of 1.43 points for waitresses while there is no evidence of statistically significant discrimination against female opticians.

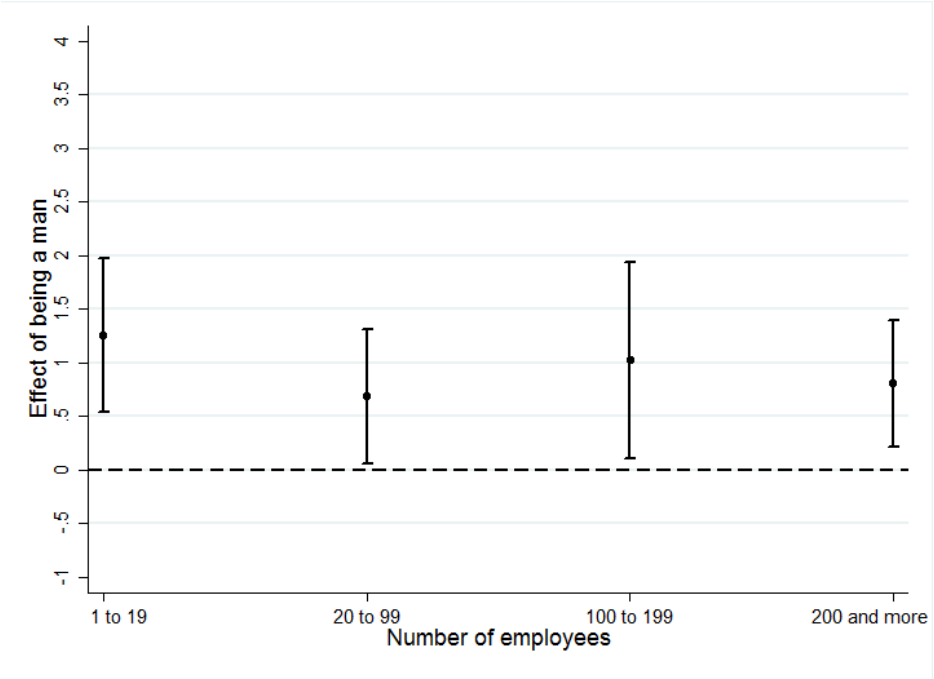
Hence we find some evidence that status moderates discrimination. In contrast to existing evidence (Neumark et al., 1996), we find more severe discrimination against women for occupations that rank lower in the status hierarchy. Note that our analysis includes 126 occupations and two different measures of status that we take from the literature compared to previous studies that were restricted to a small set of occupations and no validated measures of status. However, it should be kept in mind that we are only considering occupations that do not require a university degree. It could very well be true that the picture is reversed for high-status occupations requiring tertiary education and for senior management jobs.

4.4.4 Firm size, professionalization of the hiring procedure, and tightness of the labor market

Previous research has found a negative relationship between firm size and discrimination (Engels, 2015, Akar et al., 2014). If the size of the firm is somehow reflected in the individual evaluations made by our respondents, we would expect discrimination to be a decreasing function of firm size. We measure the size of the firm by the number of employees in the firm and create four categories of one to 19 employees, 20 to 99, 100 to 199 and 200 or more.[35]

Figure 10 shows that, controlling for the vignette characteristics, for applicants evaluated by the smallest firms (1 to 19 employees) the disadvantage of female applicants is the biggest and amounts to 1.26 points. The disadvantage is smaller for the other three categories, but it is still sizeable and statistically significant, and no linear relationship between firm size and the amount of discrimination emerges.

Figure 10: Evaluation of fictitious applicants by gender and firm size



Note: The figure presents the effect of being a man along with 95% confidence intervals obtained by a RE regression of the evaluation on a gender dummy, the firm size measured by the number of employees, the interactions between the firm size and the gender dummy and all other vignette characteristics.

When we use the number of apprenticeships in the firm instead of the number of employees as a measure of firm size, we again find no clear evidence that discrimination depends on firm size. The coefficient on the interaction term between gender and the number of apprentices is not significantly different from zero. The estimated effect of being female is -0.86 for firms with four apprentices (the median) and -0.77 for the 0.9-quantile (31 apprentices). In both cases, the gender

[35] We only count those employees for whom the firm has to make social security payments.

effect is significantly different from zero. Only for firms with a very large number of apprentices, the gender effect does no longer reach statistical significance.

A related variable potentially moderating gender discrimination is the professionalization of the hiring procedure. A more professionalized and standardized recruitment process might counteract gender discrimination, for example because standardized evaluation criteria are applied or because managers with a taste for discrimination have to justify their decisions vis-à-vis others. Of course, the evidence from our vignettes can only pick this up if individual managers who respond to our survey have internalized the features of their firms recruitment process. We measure a firm's degree of professionalization by the number of steps that a firm's recruitment process typically takes. These include, for example, the job interview, recruitment tests using work-related or school-related questions, intelligence tests or personality tests, or trial work in the firm. Clearly, not all of the steps are required in every firm, but we interpret a lack of many of these steps as a lower level of professionalization. If we interact the number of steps in the recruitment process linearly with gender, we find the coefficient on the interaction term to be -0.09 and not significantly different from zero. There is also no evidence that discrimination depends on the number of steps if we use a dummy for more than three steps or a dummy for more than four steps in the recruitment process.

Finally, we consider whether labor market conditions moderate gender discrimination. Specifically, human resource managers can be expected to discriminate less against women if they face a limited supply of applicants. As an indicator of labor market tightness, we look at the percentage of apprenticeships of an occupation in a firm that have remained unfilled in the past years (a categorical variable with the categories 0, 25% or less, 25 to 50%, and 50% or more). We find no evidence of gender discrimination if more than 50 percent of the apprenticeships have remained unfilled (effect size: -0.30), but significantly more and statistically significant discrimination if only 25 to 50 percent (effect size: -1.40), less than 25 percent (effect size: -1.38) or no apprenticeships (effect size: -0.84) have remain unfilled in the last three years. Hence, for those firms that have trouble filling their job openings, we find no indication of discrimination.

In summary, we find little evidence that the size of a firm or the degree of professionalization of the recruitment procedure have a systematic influence on discrimination at the early stage of the selection process that we are focusing on. On the other hand, we observe that human resource managers seem to discriminate less when the supply of suitable applicants is low.

4.5 Disentangling the effects of firms, occupations, and industries

The previous sections were devoted to the analysis of how discrimination varies with several firm-related or occupation-related moderator variables. In this section we attempt to determine which of these variables have explanatory power for the observed gender difference. Existing field experiments have analyzed the effect of some of these moderators alone, but it is likely that by varying,

for example, the share of female employees in an occupation, these studies also vary other moderator variables such as the status of these occupations, the salary etc. In fact, the share of currently employed female apprentices and the variables related to status (educational requirements, salaries, status indices) are weakly correlated in our sample.[36]

We consider the effects of all our variables of interest in one regression. Thus, we employ interaction terms between the gender dummy and all moderator variables analyzed in the previous subsection. We then regress the applicant's evaluation on our moderator variables and the interaction between these moderators and gender.[37] We additionally control for all vignette characteristics. For comparability, we standardize all variables such that the coefficients can be interpreted in terms of standard deviations.

In a first step, we include the variables related to status, namely the wage paid, the education required, and one of the status measures in the RE regression.[38] The results can be found in columns (1) and (2) of Table 4. Column (1) shows that when using the ISEI, the coefficients on the interaction terms between the hourly wage and gender, the education index and gender, and the ISEI and gender are all not statistically different from zero. The same holds true for the interaction terms between gender and firm size as measured by the number of apprentices, between gender and the number of steps in the recruitment process, between gender and labor market tightness, and gender and the dummies for the classes of occupations (business or administrative and educational or nursing). Hence, we find no evidence that any of these variables are moderating the gender difference in evaluations.

In contrast, we find a positive effect for the interaction between the share of female apprentices and gender. An increase in the share of currently employed female apprentices by one standard deviation leads to a decrease in the effect of being female on the evaluation by 0.84 points. This effect is highly significant. It is also relatively large, since it is bigger in absolute terms than the effect of having only an intermediate (3.4) instead of a relatively good GPA (2.8) on evaluations (not reported in the table). Quantitatively, it implies that an increase in the share of female apprentices by 50 percentage points raises female evaluations relative to male evaluations by 1.30

[36] In particular, the share of female apprentices is weakly positively correlated with the educational requirements of an occupation (r=0.27 with the typical school-leaving degree and r=0.27 using the education index). It is weakly negatively correlated with the hourly wage of an apprenticeship occupation (r=-0.17). The share of female apprentices is moderately positively correlated with the ISEI (r=0.56) and weakly positively correlated with the SIOPS (r=0.39).

[37] The moderator variables are the occupation classification of the apprenticeship as technical, business/administrative, or educational/nursing, share of women as apprentices in the firm, average hourly wage of an occupation, educational requirement of an occupation measured by the index, status index (ISEI or SIOPS), firm size measured by the number of apprentices in the firm, professionalization of the human resource management and labor market tightness measured by the ratio of applicants invited to the next step of the recruitment process to applications received. The main results are unchanged if we use the number of employers as the measure of firm size, the Classification of Economic Activities instead of the classification of apprenticeship occupations into technical, business/administrative, or educational/nursing occupations and education indicators instead of the education index.

[38] For ease of interpretation, we use the education index introduced in Appendix A.6 where a higher number stands for more demanding educational requirements.

points.

Table 4: Disentangling the effects of firms, occupations, and industries

	(1)	(2)	(3)	(4)	(5)	(6)
Fem	-0.84**	-0.80**	-0.65**	-0.86**	-0.76**	-0.73**
	(0.284)	(0.283)	(0.278)	(0.281)	(0.282)	(0.277)
Fem*Share Fem	0.84**	0.89**	0.78**	0.65**	0.65**	0.61*
	(0.286)	(0.269)	(0.241)	(0.235)	(0.240)	(0.242)
Fem*Hourly Wage	0.26	0.28	0.21			
	(0.241)	(0.219)	(0.185)			
Fem*Education	-0.06	-0.10		-0.00		
	(0.216)	(0.216)		(0.191)		
Fem*ISEI	- 0.07				0.05	
	(0.288)				(0.226)	
Fem* SIOPS		-0.04				0.11
		(0.211)				(0.196)
Fem*Firm Size	-0.21	-0.17	-0.09	-0.18	-0.09	-0.13
	(0.219)	(0.219)	(0.170)	(0.218)	(0.168)	(0.170)
Fem*Steps	-0.06	-0.05	-0.04	-0.08	-0.07	-0.06
	(0.179)	(0.180)	(0.174)	(0.179)	(0.173)	(0.175)
Fem*Tightness	-0.20	-0.19	-0.17		-0.06	-0.02
	(0.182)	(0.182)	(0.177)	(0.182)	(0.176)	(0.179)
Occupation type						
Fem*Business/administrative	-0.00	-0.01	-0.18	0.16	-0.02	-0.03
	(0.481)	(0.483)	(0.474)	(0.472)	(0.469)	(0.461)
Fem*Educational/nursing	-0.40	-0.45	-0.73	-0.27	-0.62	-0.58
	(0.746)	(0.701)	(0.689)	(0.695)	(0.734)	(0.690)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2705	2705	2874	2888	2888	2888
R-squared	0.19	0.18	0.17	0.17	0.17	0.16

Notes: ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses. "Fem" indicates a dummy for female applicants. "Share Fem" indicates the share of currently employed women in an apprenticeship occupation. Education is measured by the education index, and firm size is measured by the number of current apprenticeships in the firm. Controls are the vignette characteristics (unemployment spell, information on unemployment spell, father's occupation, mother's occupation, GPA, social behavior, number of absent days). The omitted categories are the class of technical apprenticeship occupations, age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, no absent days. The table reports only the coefficients on the interaction terms.

In column (2) we employ the SIOPS instead of the ISEI as our status measure. The results are qualitatively unchanged. Observing no statistically significant effects of the status-related variables on the gender difference in evaluations might be caused by including different variables for status that are correlated with each other. Indeed, we find that the educational requirement, the average wage in the occupation, and the status index are correlated.[39] Therefore, we re-run

[39]The wage variable is weakly to moderately correlated with educational requirements (r=0.39 with dummies for

our RE estimation including only one of the status-related variables at a time (see columns (3) to (6)). We again find no evidence that any of the status-related variables (salary, educational requirement, ISEI or SIOPS) explain the gender differences in evaluations. A detailed discussion of these results can be found in Appendix A.7.

We draw a number of conclusions from the analysis presented in this section. The difference in evaluations between male and female applicants is mainly explained by the share of female applicants in an occupation. Discrimination against female applicants is a negative function of the share of female apprentices currently employed. Further analyses based on a sample split for male and female applicants show that most of the difference is due to women being evaluated better when the share of female applicants increases while the difference is not caused by a worse evaluation of male applicants when the share of women increases.[40] We find no evidence that the variables related to status, such as the average hourly wage of an occupation, an occupation's educational requirements, and indices of occupational status, as well as firm size, the type of the occupation, or the professionalization of the recruitment process can explain a significant proportion of the observed gender difference.

5 Conclusion

We document that female job applicants are evaluated worse than male applicants by respondents from a large sample of German firms. No matter how we split our sample by observable characteristics of firms, of occupations, or of the job applicants, discrimination against women is almost always observed, albeit to a varying degree. Our findings cannot be explained by differences in the expected variance of the unobserved productivity of men and women. The observed gender differences are also unlikely to represent statistical discrimination caused by the productivity signals of the underrepresented gender being less reliable.

Gender discrimination differs between industries, with the construction industry, manufacturing and the education and public service sector displaying a significant penalty for female applicants. Also, women fare significantly worse than men both in male-dominated professions and in occupations requiring only a lower secondary school degree. There is no clear relationship between the salaries of professions and gender discrimination, but we find that with the increasing social status of an occupation (as measured by the two indices ISEI and SIOPS), the evaluations of women relative to men improve. This is also true for very tight labor markets. Finally, there is no correlation between the size of the firm and the structure of the recruitment procedure with the amount of gender discrimination.

education levels, r=0.40 with the education index) and weakly correlated with both measures of occupational status (r=0.39 with ISEI, r=0.25 with SIOPS). The correlation between educational requirements and the status indexes is weakly to moderately positive (r=0.31 (r=0.32) using indicators for education levels (the education index) and the SIOPS, r=0.42 (r=0.43) using indicators for education (the education index) and the ISEI).

[40] The regressions are not reported in the paper, but are available upon request from the authors.

When we consider the relationship between gender discrimination and all firm- and occupation-specific moderator variables simultaneously, only the gender ratio in a profession shows a significant effect on how much a firm's evaluation of an applicant depends on the gender. In particular, women suffer large penalties in male-dominated professions.

Regarding the method of factorial survey design, the responses that we have collected lead to surprisingly clear differences between the evaluations of our candidates, with the main variables such as grades and age, having the expected effects. We believe that the validity of the answers we received is due to the fact that our survey was conducted as part of a regular panel, and that we presented human resource managers with vignettes structured like a CV where such evaluations of CVs are part of the respondents' everyday work.

One unobserved variable that could determine the productivity in certain professions is physical strength. If employers expect men to be, on average, stronger than women, this could give them an advantage, for example, in the construction industry or in manufacturing where we find the largest differences in evaluations. We would like to make two comments regarding this concern. First, many of the most common professions with a strong male dominance do not require physical strength, such as a mechatronics engineer, and many similar technical occupations. Second, the amount of strength needed for certain tasks is a function of the tools developed for them. "Handle size and tool weight are designed to accommodate the size and strength of men," as stated in a report to OSHA (Occupational Safety and Health Administration) of the US Department of Labor (1999), which recommends changes necessary to allow women to enter the construction industry. While we cannot exclude that differences in physical strength can explain some of the differences in the evaluations of men and women, this would mean, however, that at least part of the productivity differential can be removed by appropriate adjustments of the workplace.

Our results clearly indicate that in male-dominated professions, employers are less likely to hire women compared to men. Thus, discrimination is likely to push women into female-dominated jobs. This can perpetuate gender imbalances and contribute to a shortage of labor supply, for example, in technical professions, especially if women anticipate the firms' responses and therefore apply at a lower rate to professions in which they are underrepresented. Given the design of our study, we can only make causal statements regarding the effect of gender on the evaluation of an applicant and not regarding the effect of the gender ratio on discrimination. However, our results are consistent with policies that aim at a sufficient number of women in jobs where they are underrepresented to remove the existing obstacles for women and reach a tipping point. Such policies could change the gender stereotypes attached to these jobs and the atmosphere at the workplace that are often seen as culprits of the self-perpetuating gender differentials.

References

[1] Advisory Committee on Construction Safety and Health (1999). Women in the Construction Workplace: Providing Equitable Safety and Health Protection. Submitted to the Occupational Safety and Health Administration (OSHA), Department of Labor, USA. Retrieved from https:// www.osha.gov/doc/accsh/haswicformal.html#ergonomics.

[2] Aigner, Dennis J., and Glen G. Cain (1977). Statistical theories of discrimination in labor markets. ILR Review 30.2 : 175–187.

[3] Akar, G., B. Balkan, and S. Tumen (2014). Overview of Firm-Size and Gender Pay Gaps in Turkey: The Role of Informal Employment.

[4] Altonji, J.G. and R.M. Blank (1999). Race and Gender in the Labor Market. In: Ashenfelter, O. and D. Card (eds.). Handbook of Labor Economics. Vol. 3, chapter 48.

[5] Authoring Group Educational Reporting (2012). Education in Germany 2012. Bielefeld: W. Bertelsmann Verlag.

[6] Azmat, and B. Petrongolo (2014). Gender and the Labor Market: What Have We Learned From Field and Lab Experiments? Labour Economics 30, 32-40.

[7] Baethge, M., H. Solga and M. Wieck (2007). Berufsbildung im Umbruch - Signale eines überfälligen Aufbruchs. Netzwerk Bildung. Berlin: Friedrich-Ebert-Stiftung.

[8] Bayard, K., Hellerstein, J., Neumark, D. and Troske, K. (2003). New evidence on sex segregation and sex differences in wages from matched employee-employer data. Journal of labor Economics, 21(4), 887–922.

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

[10] BIBB [Bundesinstitut für Berufsbildung] (2012). Datenreport zum Berufsbildungsbericht 2012, Table A.4.10.2-5. Bonn: BIBB. Online: http://datenreport.bibb.de/html/4724.htm, last access: 23.01.2015.

[11] BMBF [Bundesministerium für Bildung und Forschung] (2012). Bundesbericht Forschung und Innovation 2012, Figure 2.3.37. Online: http://www.datenportal.bmbf.de/portal/2.3.37, last access: 23.01.2015.

[12] Bohnet, I., A. van Geen, and M. Bazerman (2016). When Performance Trumps Gender Bias: Joint Versus Separate Evaluation. Management Science 62(5), 1225–1234.

[13] Booth, A. and A. Leigh (2010). Do employers discriminate by gender? A field experiment in female-dominated occupations. Economic Letters 107, 236–238.

[14] Cohen, S. L., and K.A. Bunker (1975). Subtle effects of sex role stereotypes on recruiters' hiring decisions. Journal of Applied Psychology, 60, 566–572.

[15] Dustmann, C. (2004). Parental background, secondary school track choice, and wages. Oxford Economic Papers 56.2: 209-230.

[16] Engels, S.I. (2015). The Effect of Firm Size on Gender Wage Discrimination in the Netherlands. Economics. Retrieved from http://hdl.handle.net/2105/30064.

[17] Fershtman, C. and U. Gneezy (2001). Discrimination in a Segmented Society: An Experimental Approach, Quarterly Journal of Economics 116, 351–377.

[18] Firth, Michael (1982). Sex discrimination in job opportunities for women. Sex Roles, 8 (8), 891–901.

[19] Fitzenberger, B., S. Lickleder er, and M. Zimmermann (2015). Übergänge von der allgemeinbildenden Schule in berufliche Ausbildung und Arbeitsmarkt: Die ökonomische Perspektive. Forthcoming in: Seifried, J., Seeber, S. und Ziegler, B. (Ed.): Jahrbuch der berufs- und wirtschaftspä dagogischen Forschung 2015. Schriftenreihe der Sektion Berufs-und Wirtschaftspädagogik. Opladen: Barbara Budrich.

[20] Ganzeboom, H.B.G. (2010). A New International Socio-Economic Index [ISEI] of Occupational Status for the International Standard Classification of Occupation 2008 [ISCO-08] Constructed with Data from the ISSP 2002-2007; with an Analysis of Quality of Occupational Measurement in ISSP. Paper presented at Annual Conference of International Social Survey Programme, Lisbon, May 1 2010.

[21] Ganzeboom, H.B.G. and D.J. Treiman (2017). International Stratification and Mobility File: Conversion Tools. Amsterdam: Department of Social Research Methodology, http://www.harryganzeboom.nl/ismf/index.htm. Last revision on April 7, 2017.

[22] Ganzeboom, H.B.G. and D.J. Treiman (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. Pages 159-193. In: Jürgen H.P. Hoffmeyer-Zlotnik & Christof Wolf (Eds.), Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables. New York: Kluwer Academic Press.

[23] Gerhards, C., S. Mohr and K. Troltsch (2016). BIBB Training Panel - An Establishment Panel on Training and Competence Development 2014. gwa_1.0; Research Data Center at BIBB; Bonn: Federal Institute for Vocational Education and Training. doi:10.7803/371.14.1.2.10.

[24] Heckman, J.J (1998). Detecting discrimination. The Journal of Economic Perspectives 12.2: 101–116.

[25] Heckman, J.J. and P. Siegelman (1993). The Urban Institute Audit Studies: Their Methods and Findings.

[26] Jonsson, J.O., D.B. Grusky, M. Di Carlo, Pollak, R., and M. Brinton (2009). Microclass Mobility. Social Reproduction in Four Countries. American Journal of Sociology 114 (4), 977–1036.

[27] Kaas, L. and C. Manger (2011). Ethnic Discrimination in Germany's Labor Market. German Economic Review 13, 1–20.

[28] Knoll, B., N. Riedel, and E. Schlenker (forthcoming). He's A Chip Off The Old Block: The Persistency of Occupational Choices Among Generations. LABOUR: Review of Labour Economics and Industrial Relations.

[29] Kleven, H. J., C. Landais, and J.E. Sogaard (2017). Children and gender inequality: Evidence from Denmark. Unpublished manuscript, London School Econ.

[30] Levinson, R.M. (1975). Sex Discrimination And Employment Practices: An Experiment With Unconventional Job Inquiries. Social Problems 22, 533–543.

[31] List, J. (2004). The Nature and Extent of Discrimination in the Market Place: Evidence from the Field. Quarterly Journal of Economics 119, 49–90.

[32] Muchinsky, P. M. and Harris, S. L. (1977). The effect of applicant sex and scholastic standing on the evaluation of job applicant resumes in sex-typed occupations. Journal of Vocational Behavior, 11(1), 95–108.

[33] Neumark, D., R.J. Bank and K.D. Van Nort (1996). Sex Discrimination in Hiring in the Restaurant Industry: An Audit Study. Quarterly Journal of Economics 111(3), 915–42.

[34] Neumark, D. (2012). Detecting Discrimination in Audit and Correspondence Studies. Journal of Human Resources 47(4), 1128–1157.

[35] Olian, J.D., D.P. Schwab and Y. Haberfeld (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. Organizational Behavior and Human Decision Processes, 41(2), 180–195.

[36] Petit, P. (2007). The Effects Of Age And Family Constraints On Gender Hiring Discrimination: A Field Experiment In The French Financial Sector. Labour Economics 14, 371–391.

[37] Riach, P. A. and Rich, J. (1987). Testing For Sexual Discrimination in the Labor Market. Australian Economic Papers 26, 165–178.

[38] Riach, P. A. and Rich, J. (2002). Field Experiments Of Discrimination In The Market Place. Economic Journal 112, F480–F518.

[39] Riach, P. A. and Rich, J. (2006). An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market. Advances in Economic Analysis & Policy 6(2), article 1.

[40] Schein, Virginia E. (1973). The relationship between sex role stereotypes and requisite management characteristics. Journal of Applied Psychology 57(2), 95-100.

[41] Statistisches Bundesamt (Destatis) (2016). Verdienststrukturerhebung 2014 Fachserie 16 Heft 1.

[42] Treiman, D.J. (1977). Occupational Prestige in Comparative Perspective. New York: Academic Press.

[43] Weichselbaumer, D. (2004). Is it Sex or Personality? The Impact of Sex-Stereotypes on Discrimination in Applicant Selection. Eastern Economic Journal 30, 159–86.

A Appendix

A.1 Introductory text for the vignette in English and the original German vignette

"We will now show you short descriptions of five fictitious persons. Please imagine that these persons are applying for an apprenticeship for the most frequent apprenticeship occupation at your company, either online or with a written application. All applications contain a good cover letter. All applicants are from Germany. Please evaluate how likely it is that an applicant will receive an invitation for the next step of the recruitment procedure in your company (after processing the written application) based on the information provided in the CV."

Figure 11: Example of a vignette as shown to the respondents

Geschlecht: Weiblich	**Schulbildung:** Mittlerer Schulabschluss 2012
Geburtsdatum: 03.11.1995	**Notendurchschnitt Abschlusszeugnis:** 3,4
Beruf des Vaters: Mitarbeiter in Ihrem Betrieb	**Sozialverhalten laut Zeugnisanlage:** Überwiegend gut
Beruf der Mutter: Altenpflegerin	**Unentschuldigte Fehltage:** Drei
Derzeitige Tätigkeit: Aushilfsjob	**Tätigkeit seit Schulabschluss:** Einjährige berufsvorbereitende Maßnahme und abgebrochene Ausbildung; seit Dez. 2013: Aushilfsjob

A.2 Detailed regression results

Table 5: Marginal effects of ordered probit estimation

Evaluation	Marginal effect	Standard error
1	0.065**	0.013
2	0.024**	0.005
3	0.023**	0.004
4	0.014**	0.003
5	0.008**	0.002
6	-0.002*	0.001
7	-0.012**	0.003
8	-0.029**	0.006
9	-0.025**	0.005
10	-0.067**	0.013
Observations	3164	

Notes: ** $p < 0.01$, * $p < 0.05$. The second column reports the marginal effects of being female on the probability of observing an evaluation as indicated by the first column. The marginal effects are obtained from an ordered probit estimation from the evaluation on the vignette characteristics (see Table 2). Standard errors are obtained using the Delta-method.

Table 6: Discrimination across classes of occupations

	Coefficient	Standard errors
Female*Business/administrative	-0.82*	0.376
Female*Educational/nursing	0.14	0.549
Female	-1.23	0.236
Business/administrative	-0.65*	0.263
Educational/nursing	-0.69	0.408
Age 18	-0.89**	0.085
Age 20	-0.94**	0.089
No info about gap	-0.00	0.079
Worse GPA	-0.77**	0.063
Intermediate social behavior	-0.45**	0.063
Three absent days	-0.81**	0.063
Mother elderly-care nurse	0.10	0.063
Father's occupation		
Insurance salesman	-0.03	0.084
Teacher	-0.03	0.084
Employee in firm	0.29**	0.105
Constant	8.65**	0.210
Observations	3164	
R-squared	0.13	

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the result of a RE regression. Omitted categories are technical apprenticeship occupation, age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days.

Table 7: Discrimination across classes of firms

	Coefficient	Standard errors
Female*Agriculture, Forestry & Mining	-0.66	0.925
Female*Manufacturing	-0.70	0.571
Female*Construction	-2.14*	0.841
Female*Trade & Repair	-0.41	0.668
Female*Business Services	-0.39	0.664
Female*Other Services	-0.29	0.653
Female*Public Service & Education	-0.89	0.649
Female	-0.34	0.449
Agriculture, Forestry & Mining	0.29	0.682
Manufacturing	-0.03	0.413
Construction	-0.61	0.592
Trade & Repair	-0.51	0.484
Business Services	-0.76	0.473
Other Services	-0.16	0.446
Public Service & Education	-0.58	0.451
Age 18	-0.89**	0.085
Age 20	-0.93**	0.089
No info about gap	-0.01	0.079
Worse GPA	-0.77**	0.063
Intermediate social behavior	-0.45**	0.063
Three absent days	-0.81**	0.063
Mother elderly-care nurse	0.10	0.063
Father's occupation		
Insurance salesman	-0.02	0.084
Teacher	-0.04	0.084
Employee in firm	0.29**	0.105
Constant	8.78**	0.347
Observations	3164	
R-squared	0.13	

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the result of a RE regression. Omitted categories are Medical & Nursing Services, age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days. Classification of firms based on WZ 08.

Table 8: Discrimination and variables of interest

Variable of interest	Share Female	Wage	Education	ISEI	SIOPS	Firm Size	Steps
Female*Variable of interest	1.85**	0.02	0.40	0.03*	0.03	0.00	-0.09
	(0.539)	(0.051)	(0.289)	(0.014)	(0.023)	(0.005)	(0.138)
Female	-1.66**	-1.13	-1.70**	-1.92**	-2.28*	-0.87**	-0.57
	(0.297)	(0.861)	(0.590)	(0.558)	(0.987)	(0.185)	(0.536)
Variable of interest	-0.33	-0.10**	-0.88**	-0.02*	-0.03	0.01**	-0.12
	(0.374)	(0.036)	(0.213)	(0.010)	(0.017)	(0.003)	(0.094)
Age 18	-0.88**	-0.88**	-0.91**	-0.88**	-0.88**	-0.89**	-0.89**
	(0.086)	(0.086)	(0.088)	(0.086)	(0.086)	(0.085)	(0.085)
Age 20	-0.92**	-0.93**	-0.99**	-0.92**	-0.92**	-0.94**	-0.94**
	(0.090)	(0.090)	(0.092)	(0.090)	(0.090)	(0.089)	(0.089)
No info about gap	-0.01	-0.01	0.01	-0.01	-0.01	0.00	-0.01
	(0.079)	(0.079)	(0.082)	(0.079)	(0.079)	(0.079)	(0.079)
Worse GPA	-0.77**	-0.77**	-0.76**	-0.77**	-0.77**	-0.77**	-0.76**
	(0.064)	(0.064)	(0.065)	(0.063)	(0.063)	(0.063)	(0.063)
Intermediate social behavior	-0.47**	-0.47**	-0.47**	-0.47**	-0.47**	-0.45**	-0.45**
	(0.064)	(0.064)	(0.065)	(0.063)	(0.063)	(0.063)	(0.063)
Three absent days	-0.82**	-0.82**	-0.81**	-0.82**	-0.82**	-0.81**	-0.81**
	(0.064)	(0.064)	(0.065)	(0.063)	(0.063)	(0.063)	(0.063)
Mother elderly-care nurse	0.11	0.12	0.11	0.12	0.12	0.11	0.10
	(0.064)	(0.064)	(0.065)	(0.063)	(0.063)	(0.063)	(0.063)
Father's occupation							
Insurance salesman	-0.03	-0.04	-0.01	-0.03	-0.03	-0.02	-0.02
	(0.085)	(0.085)	(0.087)	(0.084)	(0.084)	(0.084)	(0.084)
Teacher	-0.05	-0.05	-0.05	-0.05	-0.05	-0.03	-0.03
	(0.085)	(0.085)	(0.087)	(0.085)	(0.085)	(0.084)	(0.084)
Employee in firm	0.29**	0.29**	0.36**	0.28**	0.28**	0.30**	0.29**
	(0.106)	(0.106)	(0.109)	(0.106)	(0.106)	(0.105)	(0.105)
Constant	8.64**	10.19**	10.17**	9.39**	9.65**	8.34**	8.92**
	(0.248)	(0.267)	(0.456)	(0.425)	(0.712)	(0.182)	(0.391)
Observations	3106	3107	2938	3121	3121	3164	3164
R-squared	0.11	0.11	0.13	0.10	0.10	0.11	0.10

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the results of RE regressions. Education refers to the education index and firm size to the number of apprentices in a firm. Omitted categories are age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days.

Table 9: Discrimination and required education

	Coefficient	Standard Error
Female*intermediate secondary school	0.59	0.370
Female* high school	0.19	0.466
Female	–1.28**	0.321
Intermediate secondary school	-0.30	0.259
High school	-0.86*	0.340
Age 18	-0.90**	0.088
Age 20	-0.98**	0.093
No info about gap	0.01	0.082
Worse GPA	-0.76**	0.066
Intermediate social behavior	-0.47**	0.066
Three absent days	-0.82**	0.066
Mother elderly-care nurse	0.11	0.066
Father's occupation		
Insurance salesman	-0.02	0.087
Teacher	-0.04	0.087
Employee in firm	0.35**	0.109
Constant	8.82**	0.267
Observations	2923	
R-squared	0.13	

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the result of a RE regression. Omitted categories are no or lowest secondary school required, age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days.

Table 10: Discrimination and firm size measured by the number of employees

	Coefficient	Standard Error
Female* Size 2	0.57	0.484
Female* Size 3	0.24	0.592
Female* Size 4	0.46	0.474
Female	-1.26**	0.365
Size 2	0.11	0.341
Size 3	-0.18*	0.421
Size 4	-0.02	0.331
Age 18	-0.89**	0.085
Age 20	-0.93**	0.089
No info about gap	-0.01	0.079
Worse GPA	-0.77**	0.063
Intermediate social behavior	-0.45**	0.063
Three absent days	-0.81**	0.063
Mother elderly-care nurse	0.11	0.063
Father's occupation		
Insurance salesman	-0.02	0.084
Teacher	-0.02	0.084
Employee in firm	0.29**	0.105
Constant	8.49**	0.284
Observations	3164	
R-squared	0.10	

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the result of a RE regression. Firm size 2 is 20 to 99, firm size 3 is 100 to 199, firm size 4 is 200 or more employees. Omitted categories are 1 to 19 employees, age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days.

Table 11: Discrimination and professionalization of the recruitment process

	Few steps: Steps ≤ 3		Few steps: Steps ≤ 4	
	Coefficient	Standard Error	Coefficient	Standard Error
Female*Few steps	0.21	0.349	0.04	0.349
Female	-1.01**	0.258	-0.93*	0.260
Few steps	0.19	0.247	0.42	0.247
Age 18	-0.89**	0.085	-0.89**	0.085
Age 20	-0.93**	0.089	-0.93**	0.089
No info about gap	-0.01	0.079	-0.01	0.079
Worse GPA	-0.77**	0.063	-0.77**	0.063
Intermediate social behavior	-0.45**	0.063	-0.45**	0.063
Three absent days	-0.81**	0.063	-0.81**	0.063
Mother elderly-care nurse	0.10	0.063	0.10	0.063
Father's occupation				
Insurance salesman	-0.02	0.084	-0.02	0.084
Teacher	-0.03	0.084	-0.03	0.084
Employee in firm	0.29**	0.105	0.29**	0.105
Constant	8.82**	0.224	8.15**	0.224
Observations	3164		3164	
R-squared	0.10		0.10	

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the results of RE regressions. "Few steps" is an indicator for firms with less than or exactly three [four] steps in their recruitment process in column 2 and 3 [4 and 5]. Omitted categories are more than three [four] steps in column 2 and 3 [4 and 5], age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days.

Table 12: Discrimination and tightness of the labor market measured by the ratio of invitations to applications

	Coefficient	Standard Error
Female* Tightness	-0.25	0.580
Female	-0.73**	0.320
Tightness	1.44	0.420
Age 18	-0.89**	0.089
Age 20	-0.96**	0.093
No info about gap	-0.01	0.082
Worse GPA	-0.77**	0.066
Intermediate social behavior	-0.45**	0.066
Three absent days	-0.84**	0.066
Mother elderly-care nurse	0.10	0.066
Father's occupation		
Insurance salesman	-0.03	0.087
Teacher	-0.04	0.088
Employee in firm	0.32**	0.109
Constant	7.87**	0.194
Observations	2938	
R-squared	0.12	

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the result of a RE regression. Tightness is the ratio of invitations to the next step of the recruitment process to applications. Omitted categories are age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days.

Table 13: Discrimination and tightness of the labor market measured by the number of unfilled apprenticeship positions

	Coefficient	Standard Error
Female* vacancies <25%	-0.54	0.524
Female* vacancies 25% to 50%	-0.56	0.736
Female* vacancies >50%	0.55	0.729
Female	-0.84**	0.202
vacancies <25%	0.21	0.346
vacancies 25 to 50%	0.92	0.535
vacancies >50%	0.74	0.444
Age 18	-0.90**	0.086
Age 20	-0.96**	0.090
No info about gap	-0.00	0.079
Worse GPA	-0.76**	0.064
Intermediate social behavior	-0.45**	0.064
Three absent days	-0.80**	0.064
Mother elderly-care nurse	0.10	0.064
Father's occupation		
Insurance salesman	-0.01	0.084
Teacher	-0.04	0.085
Employee in firm	0.32**	0.106
Constant	8.34**	0.194
Observations	3124	
R-squared	0.11	

Notes: ** $p < 0.01$, * $p < 0.05$. The table shows the result of a RE regression. Omitted categories are no vacancies in the apprenticeship occupation, age 16, information about gap, warehouse clerk, elderly-care nurse, better GPA, good social behavior, and no absent days.

A.3 External data sources

Table 14: External data sources

Variable	Description of Data Source	Key
Share of female apprentices (main text)	Datenbank Auszubildende, Federal Institute for Vocational Education and Training (BIBB) (for apprenticeship occupations regulated by the BBiG or HwO); Berufliche Schulen 2014 Fachserie 11 Reihe 2, Federal Statistics Office (Destatis) (for other apprenticeship occupations)	KldB 2010
Average hourly wage (main text)	Verdienststrukturerhebung 2014 Fachserie 16 Heft 1, Federal Statistical Office (Destatis)	KldB 2010
International Socio-Economic Index of occupational status	International Stratification and Mobility File: Conversion Tools by Ganzeboom and Treiman	ISCO-08
Standard International Occupation Prestige Score (SIOPS)	International Stratification and Mobility File: Conversion Tools by Ganzeboom and Treiman	ISCO-08

Note: The table lists all datasets used that are not part of the Company Panel on Qualification and Competence Development of the BIBB (Federal Institute of Vocational Education and Training).

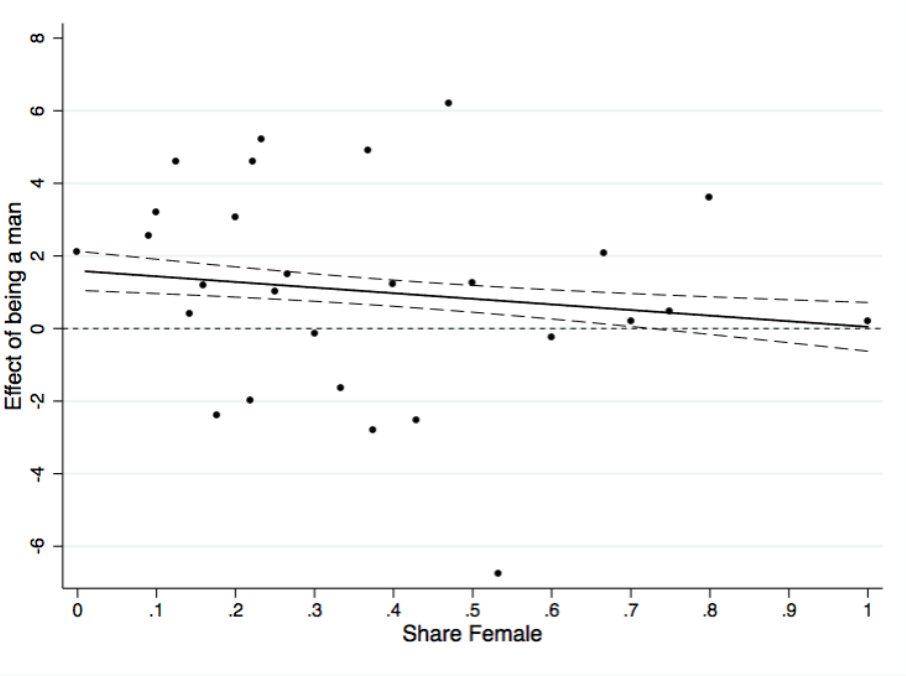
A.4 Male- versus female-dominated professions using self-reported data of the gender ratio of apprenticeships

Using the firms' self-reported shares of currently employed female apprentices, the relationship between this share and gender discrimination does not seem to be as clear-cut as for the national averages of the share of female apprentices in a profession (see Figure 12). If we regress the evaluation on a gender dummy, the share of female apprentices, the interaction term between these two variables and other vignette characteristics, the coefficient on the interaction term is significantly different from zero and amounts to 1.54 (see also Figure 12). Thus, an increase by 50 percentage points is associated with an increase in the evaluation of female applicants relative to male applicants by 0.77 points.

If we interact both the indicator for female-dominated and for male-dominated jobs with the gender dummy in one regression, the results are as follows. For gender-balanced apprenticeships we find neither evidence of discrimination against men nor against women. We again see a sizable negative effect of being female in male-dominated occupations (compared to gender-neutral occupations). We find no significant difference in discrimination between female-dominated and gender-neutral occupations.

We take away two messages from this analysis using the self-reported shares of female apprentices in each firm. First, the relationship between the self-reported share of female apprentices and gender discrimination is less clear-cut than when we use national averages. Second, our main findings about the influence of the share of female apprentices still hold.

Figure 12: Difference in evaluations between male and female applicants by share of women in apprenticeships per firm



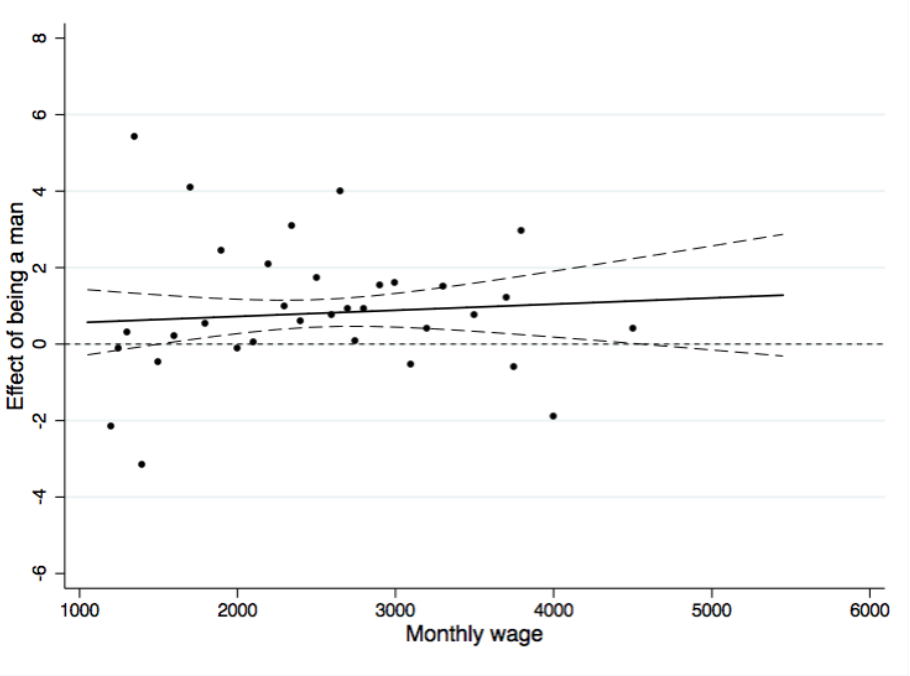
Note: The dots represent gender differences in the evaluation of applicants. The solid line shows the effect of being a man obtained by a RE regression of the evaluation on a gender dummy, the self-reported share of women currently employed in an apprenticeship occupation, the interaction between the share of women and the gender dummy and all other vignette characteristics. The dashed lines indicate 95% confidence intervals for this effect.

A.5 High- versus low-wage occupations using self-reported wage data

The dataset contains information on the average monthly wages for workers in a firm. Based on this we construct two measures of (occupational) wages in order to relate the wage level to discrimination. Note that respondents report monthly wages for all occupations in the firm and do not report the specific wages for the apprenticeship occupation we analyze. They report wages separately for unskilled ("einfache Tätigkeit"), skilled ("qualifizierte Tätigkeit"), and highly-skilled work ("hochqualifizierte Tätigkeit"). As our first measure we use the wages reported for employees with skilled jobs ("qualifizierte Tätigkeit"), since these are typically the jobs performed by employees who have completed an apprenticeship. We use the midpoints of each of the three wage categories for the respondents for which we only observe the categorized monthly wage. Monthly wages then range from 1,050 Euro to 5,700 Euro. The mean monthly wage is 2,516 Euro.

Figure 13 displays how the difference in evaluations between men and women varies with the monthly wage. The graphical evidence suggests that there might a negative relationship between the difference in evaluations between male and female applicants and wage payments. In our typical RE estimation, however, the coefficient on the interaction term is not significantly different from zero.

Figure 13: Difference in evaluations between male and female applicants by monthly wage of occupation



Note: The dots represent gender differences in the evaluation of applicants. The solid line shows the effect of being a man obtained by a RE regression of the evaluation on a gender dummy, the monthly wage of an occupation, the interaction between the monthly wage of an occupation and the gender dummy and all other vignette characteristics. The dashed lines indicate 95% confidence intervals for this effect.

As our second measure, we construct an index that summarizes the three wage categories reported (for unskilled, skilled and highly skilled jobs) to one composite measure of the wage level in the respective firm.[41] This index can take on values between 1 and 6. The index takes on the value 1 if a firm pays wages within the lowest wage category for unskilled, skilled and highly-skilled work (i.e., below 1,000 Euro). It takes on the value 6, if a firm pays wages within the highest wage category for all three types of labor (i.e., 4,500 Euro or more).

The random-effects regression delivers a coefficient on the interaction term that is -0.09,but that is not statistically significant. But we find that the gender effect is not significantly different from zero for low wages while it gains statistical significance at an index value of 2.5.

In summary, we find no strong evidence of a relationship between gender discrimination and the self-reported monthly wages of a firm. If at all, higher wages are accompanied by more discrimination against women.

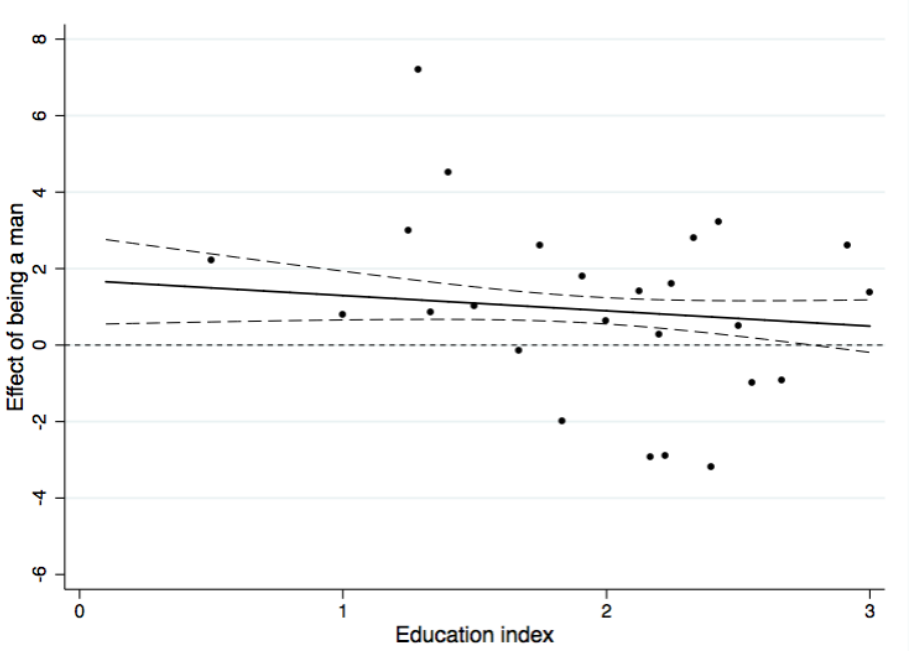
[41] Our index is given by the sum of the wage level for unskilled workers (if reported) plus the wage level for skilled workers (if reported) plus the wage level for highly-skilled workers (if reported) relative to the number of skill levels the monthly wage is reported for.

A.6 Educational requirements using an index

Instead of considering the most frequent school-leaving qualification, we now construct an index based on the firms' self-reported educational requirements. The index is equal to 3 if all apprentices of the firm in the apprenticeship occupation completed high school, it is equal to 0 if all apprentices of the firm in the apprenticeship occupation have no school-leaving qualification, and it takes on values between 0 and 3 depending on the number of apprentices with each school-leaving qualification relative to the total number of apprentices.[42] This measure of educational requirements complements our previous measure in that it focuses on differences between firms with respect to their overall cohorts of apprentices.

Figure 14 plots the average differences in evaluations against the education index. The gender difference in the evaluations seems to be most pronounced for firms with lower educational requirements for their apprentices.

Figure 14: Gender difference in evaluations by education index



Note: The dots represent gender differences in the evaluation of applicants. The solid line shows the effect of being a man obtained by a RE regression of the evaluation on a gender dummy, the education index, the interaction between the education index and the gender dummy and all other vignette characteristics. The dashed lines indicate 95% confidence intervals for this effect.

Controlling for the vignette characteristics, the correlation between discrimination against women and the education level (as measured by the education index) is statistically not significantly different from zero. The coefficient on the gender-education interaction is positive, indicating that

[42]The index is given by (number of apprentices with no school-leaving qualification*0+number of apprentices who completed secondary school only*1+number of apprentices who completed intermediary secondary school*2+number of apprentices who completed high school*3)/total number of apprentices.

higher educational requirements are accompanied by a smaller difference in evaluations between male and female applicants, but it is insignificant. At the mean (median) of the index, there is statistically and economically significant discrimination amounting to 0.93 (0.91) points. Only when the index approaches three, which means that almost all apprentices completed high school, is the discrimination against women 0.52 and statistically not distinguishable from zero anymore.

A.7 Can salaries, educational requirements, and status explain discrimination?

In order to investigate more thoroughly whether none of the status-related variables has explanatory power, we run a number of regressions where we only use one of the variables related to status.

First, we include the average hourly wage along with the other moderator variables (the share of female apprentices, the occupation type, firm size, the number of steps in the recruitment process), the interaction between the hourly wage and gender, the interaction between gender and the moderators and other control variables in the RE regression (Table 4, column (3)). The coefficient on the wage-gender interaction is statistically insignificant. For the other moderator variables the results are unchanged. There is a statistically significant and positive influence of the share of current female apprentices on the effect of being a female applicant. Increasing the share of female apprentices by one standard deviation (0.32 percentage points) is associated with an improvement in the evaluations of female applicants relative to the evaluation of male applicants by 0.65 points. There is no statistically significant influence on the effect of being female on evaluations by firm size, degree of professionalization nor type of occupation.

Second, we replace the hourly wage by the education index and introduce an interaction term for the educational requirement and gender (Table 4, column (4)). The education-gender interaction term is economically and statistically not significant. The coefficient on the interaction term between gender and the share of female apprentices increases slightly and remains highly statistically significant and sizeable.

Third, we subsequently include the ISEI and the SIOPS (Table 4, column (5) and column (6). Again, we find no evidence that status (as measured by the ISEI or SIOPS), firm size, occupation type, or degree of professionalization influence gender-differences. The coefficient on the interaction term between gender and the share of female apprentices decreases slightly, but we still find a strong effect of the share of currently employed female apprentices on the gender dummy.

Discussion Papers of the Research Area Markets and Choice 2017

Research Unit: **Market Behavior**

Research Unit: **Economics of Change**

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice