Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

When does gender occupational segregation start? An experimental evaluation of the effects of gender and parental occupation in the apprenticeship labor market[☆]

Ana Fernandes [a,*,1,2], Martin Huber [b,2], Camila Plaza [c,2]

[a] *Bern University of Applied Sciences, Institute for New Work, Brückenstr. 73, CH-3005 Bern, Switzerland*
[b] *University of Fribourg, Department of Economics, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland*
[c] *Statista, Johannes-Brahms-Platz 1, DE-20355 Hamburg, Germany*

ARTICLE INFO

ABSTRACT

The apprenticeship market is the earliest possible entry point into the workforce in developed economies. Since early labor market shocks are likely magnified throughout professional life, avoiding mismatches between talent and occupations – for example due to gender- or status-based discrimination – appears crucial. This experimental study investigates the effects of applicant gender and its interaction with parental occupation on the probability of receiving an invitation to an interview in the Swiss apprenticeship labor market. We find no robust evidence of differential treatment by employers in most cases. Policies aimed at fostering gender equality across occupations should therefore focus on removing gender related educational or cultural barriers influencing occupational choices at young ages.

1. Introduction

This paper presents a so-called correspondence test based on experimentally sending out fictitious applications to vacancies in the Swiss apprenticeship labor market in order to assess the effects of applicant gender and its interaction with parental occupation on employers' callback rates. Our paper thus aims to answer two questions. The first is whether employers are making gender stereotypically oriented decisions already at the earliest point of entry into the labor market, i.e. apprenticeships. Second, we explore whether the parental occupation of applicants affects the labor market chances of their offspring. By and large, we do not find robust statistical evidence for differential treatment by employers in terms of callbacks (that is, invitations to

an interview, assessment center, or to a trial apprenticeship) with one noticeable exception.

One major motivation for our study is the empirically observed gender occupational segregation between males and females (see e.g. Cortes & Pan, 2018 for a recent overview of evidence and preference-based explanations of gendered occupational choice). Because this phenomenon is associated with less favorable labor market outcomes for women – as wages in female-dominated professions tend to be lower than wages in male-dominated ones (Blau & Kahn, 1996) – its causes are the object of intense scrutiny. The experimental literature (e.g. correspondence testing) has attempted to uncover evidence of potential demand-side effects. Employers would contribute to gender

occupational segregation if they preferably hired women for female-dominated occupations and, vice-versa, men for male-dominated occupations. This is indeed the pattern of results obtained from correspondence testing (see e.g. Rich, 2014). Becker S. O., Fernandes, and Weichselbaumer (2019), a correspondence test which included the Swiss labor market, is a recent example of this pattern as adult women experienced a much higher callback rate relative to men in female-dominated secretarial and accounting positions. Further, Kuhn, Shen, and Zhang (2020) not only show the correlation of employer explicit requests for one gender in job advertisements with the distribution of the applicant pool but they also find penalties in the callback rates for gender-mismatched applications, penalties which are higher for female applicants.

In preventing the best match between talent and occupations, demand-side effects are inefficient in addition to being socially unjust. Furthermore, differences in initial conditions in the labor market may matter more for lifetime inequality than do shocks afterwards (see Huggett, Ventura, & Yaron, 2011). Therefore, an important question is whether or not such stereotypical decisions are already present at early stages of labor market participation. While the empirical evidence described above applies to adults, it is the aim of this study to advance research by examining demand-side effects on gender occupational segregation in the apprenticeship market, the earliest point of entry into the labor market in developed economies. For this reason, we experimentally assess how applicant gender affects callback rates in the Swiss apprenticeship market.

In Switzerland, job applications typically contain very detailed personal information, including a photo and demographic details such as age and marital status, among other elements. Apprenticeship applicants are typically 14 or 15 years of age. Because of their youth, they usually do not yet have that much to say about themselves in their CVs. However, they routinely indicate the profession of their parents. This quite unique feature of the Swiss apprenticeship market allows us to investigate whether parental background affects the labor market chances of offspring, and differently so across applicant gender.

Estimates of the intergenerational persistence of income across generations of a given family (measured for example by the correlation between measures of parent and child's permanent incomes) vary across countries (see e.g. Blanden (2013) for an international comparison), with the US exhibiting greater persistence and Denmark and Sweden taking place at the low end of international estimates. The channels leading to this phenomenon are varied and complex and are often put into one of the two categories of "nature" or "nurture." Nurture-type channels could operate through parental investments in the education of their children (educational levels of parents and children are positively correlated in the data) or through parental professional networks and connections. Whatever the true underlying channels that connect incomes of parents and those of their children, the fact that incomes are positively correlated across different generations of the same family is suggestive of a higher earnings potential in a child of wealthier parents. Intergenerational income persistence may therefore lead employers to take parental background into consideration when examining apprenticeship applications.

To assess whether employers take applicant gender and parental occupation into consideration, we sent out approximately 3000 fictitious applications (containing CVs and educational certificates) via e-mail to open apprenticeship positions across four regions in Switzerland (Basel, Bern, Lausanne, and Zurich) between August and October 2018. In the applications, we randomized demographic characteristics like gender and parental occupation in order to investigate the impact on callback rates by employers, namely invitations to interviews, assessment centers, or trial apprenticeships. The employers' responses to our applicants were recorded up to February 2019. Using applications that signaled a comparable level of productivity and differed only w.r.t. the applicant's gender and/or parental occupation was key for investigating

whether employers systematically differed in their treatment of groups with particular demographics.

By and large, we find no statistical evidence for discrimination based on applicant gender in the total sample. To put this finding into perspective, our power analysis suggests that we can detect a gender effect on the call back rate which is as small as 5 percentage points with a probability slightly higher than 80%. Importantly, the absence of statistically significantly differential callback rates by gender persists once we distinguish between female- or male-dominated (or neutral) occupations.

The absence of statistically significant gender bias in employers' callback rates, in particular considering female dominated occupations, goes against well-established regularities in the experimental literature, as discussed above, although those findings pertain to the labor market of adult persons for the most part. Recent evidence for Switzerland in Becker S. O. et al. (2019) is a case in point, as mentioned. Employers' hiring decisions concerning apprenticeship positions could differ since, for this age group, potential fertility concerns are irrelevant. However, the findings for adult persons suggest that employers make hiring decisions based on stereotypical views of the candidates in relation to the gender dominating particular occupations — with fertility issues possibly lowering the callback probability of females in fertile age across occupations.

We are aware of two other studies of the apprenticeship labor market, both vignette studies unlike the present correspondence test. Kübler, Schmid, and Stüber (2018) focused on Germany whereas Fossati, Wilson, and Bonoli (2020), like ours, studied the Swiss apprenticeship labor market. The results in Kübler et al. (2018) mimic the stereotypical findings for adults. Those in Fossati et al. (2020) align with ours in that, on average, there is no gender difference in the hiring intentions of employers (although, lacking controls for the gender dominance of occupations, it is not possible to find out whether an average null effect masks stereotypical hiring decisions along the gender dominance of occupations).

We attribute the different results on gender discrimination for Germany and Switzerland in the apprenticeship labor market as originating in asymmetric labor market regulations across these two countries, as follows. The degree of employment protection in the German labor market is significantly higher compared to the Swiss case. According to the OECD Indicators of Employment Protection,[3] workers in Germany enjoy significantly greater employment protection compared to those in Switzerland. These differences provide rather different incentives for firms to train and, importantly, to *retain* their apprentices (Muehlemann, Pfeifer, Walden, Wenzelmann, & Wolter, 2010). Indeed, the apprentice retention rate by German firms is almost twice as high as that of Swiss firms, 59% versus 35.5%, see Mühlemann (2016). If retaining apprentices helps firms reduce their hiring and firing costs, we would expect them to regard their apprentices as potential future permanent employees. And for this to happen more so in the country with the more stringent employment protection. Were firms do consider their apprentices as akin to future permanent employees, one would expect the biases that characterize the labor market of adult persons to be moved forward in time to the evaluation of apprenticeship candidates. This appears to indeed be the case in Germany (Kübler et al., 2018) but not so in Switzerland, where firing (and thus hiring) is much less costly to employers.

We further do not find effects of parental occupation on callback rates in general: differences in callback rates are not statistically significant at any conventional level when accounting for multiple hypothesis

---

[3] See e.g. OECD Indicators of Employment Protection, https://www.oecd.org/els/emp/oecdindicatorsofemploymentprotection.htm, consulted online on 23 March 2020. Regarding measures for the Strictness of Employment Protection – Individual Dismissals (regular contracts), the corresponding indicator for Germany is 2.6 whereas it is only 1.6 for Switzerland (for an OECD average of 2.0). Data are for 2013, the latest available year.

testing. The one exception is stating the father's occupation to be a university professor, which boosts callbacks in a statistically and economically significant way for female applicants, but not for males. Point estimates across subsamples suggest that the aforementioned professor effect for female applications is, to a larger extent, driven (i) by the German rather than the French speaking sample, (ii) by less demanding apprenticeships from the point of view of required qualifications, (iii) by more female- rather than male-dominated apprenticeships, and (iv) by smaller rather than larger employers in terms of the number of employees. However, due to low statistical power and issues related to multiple hypotheses testing, we abstain from putting strong interpretations on the effect heterogeneities found across subsamples. The findings across subgroups generally back those of the main analysis. Specifically, when excluding the empirically rare case of having a professor as parent from our sample, we find no statistically significantly differential callback rates across gender, not even when looking into groups of occupations with a clear gender dominance pattern or type (e.g. female- or male-dominated). Importantly, the asymmetric impact of the parental profession on girls relative to boys is not driven by the chances of having a father who is a university professor being much higher for boys compared to girls as highly educated fathers are unlikely outcomes for both genders in reality, as documented below.

Going back to the intergenerational persistence of income, the literature points to stronger correlations between the earnings of parents and sons relative to those between parents and daughters (see e.g. Chadwick & Solon, 2002 or Hirvonen, 2008). In a different setting, Dohmen et al. (2011) documented a positive relationship between average occupational earnings and occupational earnings volatility. They also find that women are significantly less willing to take risk than men. This lower correlation between parental and daughter's incomes could therefore reflect the possibility that women pursue lower but more income-stable careers. Maternity and the associated child penalties (reductions in yearly earnings following the birth of a first child) would further support this possibility (see e.g. Kleven, Landais, & Søgaard, 2019). Using a long panel of administrative data, Fernandes and Hevenstone (2023) document both extremely high values for the average child penalty experienced by Swiss moms in an international context as well a clear difference in the magnitude of those penalties along the language and cultural border separating German speaking from Italian or French speaking regions within the country. Child penalties are about double in the Swiss German speaking regions relative to the remainder of the country, be it for full- or part-time working moms. Therefore, when applying to an apprenticeship position (particularly in the German speaking region of Switzerland), girls would be associated with lower future income prospects compared to boys, on average. Further, the higher intergenerational correlation of parental income with that of sons would suggest that any networking possibilities with the professor as a father that may be of interest for the company would be relatively more useful in connection with boys than with girls. Expectations of relatively higher potential productivity for daughters of professors compared to sons, therefore, do not appear to be driving the gender asymmetric effects of having a professor as a father.

An alternative explanation could be the following. If females are assumed to have a less important labor market attachment – and particularly so in the German speaking part of Switzerland – while at the same time household income is presumably high when the father is a professor, employers may perceive female applications to an apprenticeship as a strong sign of commitment to work in itself. This commitment signal might in turn be rewarded by employers. However, we need to clearly point out that we are in no position to test this conjecture. The "commitment" signal from female applicants with professors as fathers, as opposed to the more meaningful productivity signal from the male applicants (as argued before), could be one potential explanation for the disproportionate response received by girls who report such a parental background.

Concerning parental background, our results therefore provide some support for a blind recruitment procedure. Personal attributes (such parental occupation) should not be communicated to the employer in the first round of an application process in order to prevent signaling effects and set the callback chances of all applicants on an equal footing. From a policy point of view, and answering the questions outlined above, our findings represent rather good news: when disregarding the less empirically plausible case of the professor as a father, employers do not appear to gender-discriminate applicants for apprenticeship positions in the Swiss labor market — at least not to a level that we can statistically detect. Gender occupational segregation at the apprenticeship level, therefore, appears to have its roots in the occupational choice of young persons. Fostering occupational diversity thus requires removing educational or cultural barriers currently narrowing the horizons of young persons at the time of their labor market entry.

Our paper is structured as follows. Section 2 reviews the literature on labor market discrimination and correspondence testing. Section 3 provides institutional background information on the Swiss educational system and apprenticeship market. Section 4 outlines the experimental design. Section 5 provides descriptive statistics for our data while Section 6 presents the empirical results. Section 7 concludes.

2. Literature review

Our paper is closely related to the experimental literature aiming at causally assessing the prevalence of discriminatory practices. In economics, asymmetric labor market treatment of individuals for reasons unrelated to their productivity amounts to discrimination. The two main reasons for employers to discriminate offered in the literature originate from tastes (Becker, 1957), e.g. when employers or customers dislike working with a particular group in the population, or in uncertainty about the true productivity of the candidate employee (Arrow, 1973 and Phelps, 1972). The former is commonly known as taste-based discrimination and the latter as statistical discrimination.

The preference for one gender over the other as a function of occupation type could have elements of both taste-based and statistical discrimination. Employers may have a preference for candidates with the gender that matches the sex typically expected or encountered in a particular occupation, possibly reflecting stereotypical preference biases. They may also believe that such a gender-based matching is relevant for productivity (see Goldin, 2015 and Weichselbaumer, 2004 for a detailed discussion on this matter). An interesting aspect of our experiment is that, due to the young age of apprenticeship applicants, statistical discrimination against females due to fertility concerns appears less likely than for older age groups.

Field experiments (i.e. so-called audit studies and correspondence testing) are experimental methods of data collection which involve sending fictitious applications in response to real job advertisements. In correspondence testing, for example, applications including CVs that are matched in all relevant qualifications, like schooling and job experience, but which differ w.r.t. the demographic characteristics of interest (e.g. gender, ethnicity, age), are sent out in response to job advertisements. If all productivity-related characteristics are comparable, any statistically significant differences in the response rate of employers related to the demographics is indicative of discrimination. Experimental methods gained notoriety as they were able to overcome important empirical limitations of previous tools, such as omitted variables bias (see Bertrand & Duflo, 2017 and Guryan & Charles, 2013 for a discussion). The latest developments in this extensive literature have been systematized in recent surveys (Baert, 2018; Bertrand & Duflo, 2017; Neumark, 2018, and Rich, 2014).

Regarding gender discrimination, the evidence summarized in Rich (2014) and Riach and Rich (2002) suggests that women are discriminated against in male-dominated jobs and, vice-versa, men are discriminated in female dominated occupations — while such results are frequently not found for occupations lacking a clear gender pattern.

Table 1 in Carlsson (2011) provides a useful summary of results from existing studies to date. For those professions with a clear gender pattern (e.g. network technicians or secretaries), the estimated differences in callback rates between males and females are large and economically meaningful (for example, 15pp in favor of men for the former occupation, and 10 and 24pp for two different studies on secretaries, now in favor of women).[4] Becker S. O. et al. (2019), who examined secretarial and accounting positions in the German speaking labor market, find a 7.8pp difference in the average callback rates of females relative to males in Switzerland (significant at the 1% level). Thus, the aforementioned pattern of gender occupational segregation appears to go hand in hand with quantitatively marked differences in the callback rates of males and females.

While most studies consider prime aged workers, such that statistical discrimination related to family obligations could partly explain gender differences in callback rates, two vignette studies focused on the apprenticeship market as we do. Kübler et al. (2018) examined the German apprenticeship market while Fossati et al. (2020) focussed on the Swiss apprenticeship market. Kübler et al. (2018) embedded a vignette study in a nationally representative survey of German firms hiring apprentices and found females to be evaluated worse than males, on average. In line with the broad patterns described above, the female disadvantage in average callback rates disappears once the share of women in occupations is considered. Fossati et al. (2020) contact employers hiring for apprenticeship positions and ask them to evaluate potential candidates. Their focus is on the question of whether or not employers rely on productivity unrelated information when hiring, and especially so in situations of high uncertainty. Examples of those situations would be cases where some of the academic results point in different directions (for example, when the candidate has a low grade average but a very high score in an independent general test often used by employers to compare candidates outside the schooling system). Their benchmark results (Table 2) align well with ours in that gender is not significant on average and the same mostly holds for parental background. Indeed, employers react to those demographic indicators only in a very restricted subset of cases when facing uncertainty in the applicants' profiles (Table 3 and A1).

Sharing a common focus on the apprenticeship labor market as Fossati et al. (2020) and Kübler et al. (2018), our paper employs a different methodology and is the first correspondence test on the Swiss apprenticeship market. Correspondence testing poses one advantage over vignette studies in that it examines real decisions of employers faced with credible job/apprenticeship applications. Despite this formal difference, the data suggest that employers react differently to gender cues in these two neighboring countries. As mentioned in the Introduction, we rationalize this difference under the light of labor market regulations concerning employment protection.

The variation of callback rates by gender in line with typical gender dominance in given occupations would not, *a priori*, suggest the average effect (across all occupations) to be positive or negative. Indeed, a favorable effect for women in female dominated occupations could average out with an unfavorable effect in those that are male dominated. In this respect, the lack of an average gender effect on callbacks/employer appraisal from our study and Fossati et al. (2020)'s is not necessarily at odds with the experimental evidence. What stands out in our results relative to the literature, is that we find no gender differences in callback rates even when we examine callbacks within homogeneous groups in terms of gender dominance.

An extensive literature going back to Becker and Tomes (1979) and Becker and Tomes (1986) has focused on how the educational attainment – and other later-in-life outcomes – of children relate to the education of their parents as well as to the investments parents made in offspring education. Parents influence their children's outcomes through genetic as well as educational channels, which the literature has labeled as the "nature versus nurture" dichotomy. One of the goals of this literature is measuring intergenerational income persistence, thus the degree to which offspring income is related to parental income.[5] Our experimental setup fits into this literature by allowing us to measure the social status effect of parental background on the employer's selection of an apprenticeship candidate, an instance of "nurture" type effects. The measurement of such effects in an experiment is usually difficult to implement as adult persons normally do not mention parental background in their CVs. However, as described earlier, in the Swiss apprenticeship labor market, young people in their apprenticeship applications typically do.[6]

3. The Swiss education system and vocational education

In Switzerland, the constitution broadly defines the general foundations of the educational system, like obligatory free access to primary schooling. However, the core responsibilities in providing education rest with the country's 26 cantons (regional administrative units). For this reason, there is considerable variation in school systems across cantons, although there are also attempts to harmonize key aspects of compulsory schooling through the so-called HarmoS concordate. According to the State Secretariat for Education, Research, and Innovation (2013), the vast majority of students in compulsory education attend public schools, only 5% went to private schools in the academic year of 2012/2013.

According to the Swiss Coordination Centre for Research in Education (2018), compulsory schooling consists of 11 years of education in most cantons (in particular those participating in the HarmoS concordate), including two years of kindergarten attendance that starts at the age of four. After kindergarten, primary schooling typically consists of six years and lower secondary schooling of three. In the last year of primary school, students are assessed and subsequently placed into different tracks of lower secondary education that differ in terms of qualifications. Concretely, this means that classes in grades seven to nine will be taught with varying levels of difficulty, matching the different educational tracks in which students are placed. After finishing lower secondary education, and depending on the qualifications obtained, students follow one of two possibilities. They enter either the vocational education and training (VET) track, typically consisting of a dual apprenticeship system of formal education and training in a company, or the academic track, by attending either a general or specialized high school that prepares students for tertiary education, see the State Secretariat for Education, Research, and Innovation (2018).[7]

---

[5] See e.g. Lee and Seshadri (2019) for a version of the Becker–Tomes model allowing for multiple investments stages in children's education calibrated to US data, and Björklund, Jäntti, and Nybom (2017) for an empirical comparison of factors influencing intergenerational income mobility between the UK and Sweden.

[6] The more recent correspondence testing study by Bourabain, Verhaeghe, and Stevens (2020) is also a relevant reference point for our results. The authors focus on how access to Kindergarten in Belgium is affected by subtle discriminatory strategies along ethnic and social class lines. They found that middle-class parents were offered visiting times during the day, whereas working-class parents were asked to come in the evenings, if at all. Furthermore, ethnicity also played a significant role, with schools being more open to labor class Belgians than their non-Belgian counterparts. This kind of educational gatekeeping may affect the children's education and their later in life outcomes.

[7] Students typically receive career counseling concerning their professional interests and options at the age of 14. If they choose the VET pathway, then starting the apprenticeship application process is encouraged. At the age of 15 to 16, when students have accomplished compulsory education, they typically start their apprenticeships.

The different educational tracks and corresponding lower secondary education qualification levels give rise to three different "tiers" in which occupations in VET are classified. For example, students who wish to apply for a beauty apprenticeship (a 3 year program) are required to have a *Realschulabschluss*, a school degree with comparably lower qualifications. Apprenticeships with similar requirements (e.g., baker, gardener, etc.) constitute the first tier. A *Sekundarschulabschluss*, i.e. a school degree with comparably higher qualifications, possibly combined with a standardized aptitude test, is a typical requirement for second-tier occupations, such as electric technician, mechanic or dental assistant, among others. Lastly, apprenticeships in areas such as informatics or polymechanics make up our third tier and typically require a higher level degree with reasonable grades in math and physics. Please refer to Table 10 in Appendix A for the complete list of occupations targeted in our study and the corresponding tier classification.

In Switzerland, roughly two thirds of all students with completed compulsory education enter the VET track and have around 230 occupations to choose from (State Secretariat for Education, Research, and Innovation, 2018). Apprenticeships typically take between two to four years, as discussed in Kuhn, Schweri, and Wolter (2019). Most popular are dual apprenticeship programs, which combine classes at a vocational school with on-the-job training at a host company. Apprentices are employed and paid a salary which increases with each completed year. However, also (full-time) school-based VET programs exist. They are less common overall, but relatively more popular in the French and Italian speaking regions of Switzerland.

Upon successful completion of the program, apprentices receive a federal VET diploma which not only serves as recognized occupational qualification but is also a precondition for further education and higher qualifications in the chosen occupation. According to the State Secretariat for Education, Research, and Innovation (2018), the VET system is managed as a public–private partnership, with the federal and cantonal governments as well as the employers and professional organizations jointly defining the curricula, skill sets, and standards for occupations. Moreover, the employers cover the costs for on-the-job-training, salaries, and in-house courses. The cantons, on the other hand, fund the vocational schools and career guidance services.

4. Experimental design

Our correspondence test in the Swiss apprenticeship market consisted of a preparatory phase, from October 2017 to July 2018, an experimental phase, from August 2018 until February 2019, and the debriefing of the employers in March 2019.

Preparatory phase In the preparatory phase, we developed all materials required for the production of fictitious applications to open apprenticeships. We first screened apprenticeship advertisements online to learn which documents were required in the application process.[8] Furthermore, we consulted teenagers applying for apprenticeship positions in order to learn how typical applications look like. In addition, we collected CVs and motivation letters (through personal contacts as well as online sources) to use them as templates for our fictitious applications. We also prepared electronic versions (i.e. in pdf format) of school certificates for the fictitious candidates. In order to compare candidates beyond their school credentials, employers may require apprenticeship applicants to take an aptitude test. Whether or not testing is common generally depends on the occupation (with most companies hiring in a given occupation either requesting or not requesting the test results). We thus prepared electronic versions of aptitude test certificates for the fictitious candidates as well. In Appendix E, we provide further details

on how the applications were constructed as well as an example of an actual application sent.

A further task was to classify apprenticeship types w.r.t. their relative empirical importance among females and males. We relied on information about the relative popularity of specific occupations across gender provided online by the Educational Office of the Canton of Bern (*Kanton Bern, Erziehungsdirektion*) and the Office for Equality of Males and Females of the Canton of Zurich (*Kanton Zürich, Fachstelle für Gleichstellung von Frau und Mann*).[9] These classifications were further cross-checked with additional online resources on the apprenticeship market.[10] Using these criteria, we categorized occupations into clearly male-dominated, female-dominated, and (more or less) gender neutral types.

As mentioned above, a second classification concerned the level of qualifications attained in terms of lower secondary schooling. We classified apprenticeship types into three levels of requirements (or tiers) and adapted school certificates and aptitude tests accordingly to make applications look coherent concerning the skills typically expected. For the first tier, which was lowest in terms of requirements, applications contained school certificates reflecting a lower level degree (*Realschulabschluss*) and comparably low scores in the aptitude test, if the latter was required at all. For the second tier, certificates reflecting a higher level degree (*Sekundarschulabschluss*) along with intermediate grades and aptitude test scores were used. For the most demanding third tier, certificates reflecting a higher level degree along with comparably good grades and test scores were included in the application documents.

In total, 30 occupations were selected and included in the experiment, eight of which are gender neutral (e.g. baker, cook, sales assistant, designer), six female-dominated (e.g. hair dresser, dental assistant, medical practice assistant), and 16 male-dominated (e.g. gardener, carpenter, car mechanic, mason, electrician). In selecting these occupations, we took into consideration the need of having sufficiently many observations for all three gender types in the sample, guided by online search-based estimates of how many advertisements would be posted for each type. Please refer to Table 10 in Appendix A for a complete list of occupations considered, together with their gender-type and tier.

Aiming to find an acceptable balance between expected sample sizes and organizational burden in preparing and managing applications, we decided to focus on three agglomerations in the Swiss German-speaking region, namely Basel, Bern, and Zurich, and on one agglomeration in the French-speaking region, Lausanne. We prepared fictitious motivation letters, CVs, school certificates, and aptitude tests as well as two female and male profiles for either language region with varying names, addresses, and photos. Concerning names, we took the most popular choices for first names for either gender in 2004 in the German and French speaking parts, respectively, while the last names corresponded to the most frequent occurrences in the phone book in either language region.

We also picked residential addresses in the four agglomerations for the fictitious candidates. Preparing school certificates that matched these addresses turned out to be more complicated than initially expected. This was so first, because certificates look different in each

---

[8] Such information is, for instance, provided on the websites https://www.berufsberatung.ch and https://www.yousty.ch, which we accessed in late 2017.

[9] See https://www.erz.be.ch/erz/de/index/berufsbildung/grundbildung/kennzahlen_berufsbildung/kennzahlen_berufsbildung2.html and https://ffg.zh.ch/internet/justiz_inneres/ffg/de/bildung/berufswahl/_jcr_content/contentPar/morethemes/morethemesitems/factsheet_die_belieb.spooler.download.1393238737874.pdf/FFG_2013_factsheet_die_beliebtesten_berufe_von_maedchen_und_jungen.pdf, respectively, both accessed in the beginning of 2018.

[10] See for instance the following list of the 10 most popular apprenticeships for females and males in 2015: https://blog.100000jobs.ch/de/2016/09/die-top-10-der-beliebtesten-lehrstellen/, accessed in the beginning of 2018.

canton (and even over time) and, second, because of adapting certificates to the qualifications appropriate for the three different tiers of apprenticeships. While applicant addresses and school certificates match in terms of cantonal congruence for Bern and Zurich, this is not the case for Basel and Lausanne. For the latter two agglomerations, and to ensure that the whole application was coherent, we included statements in the cover letter indicating that our fictitious applicants had recently moved from a different region.

Experimental phase We aimed at sending out two applications per open apprenticeship position and to only consider one apprenticeship per employer to avoid straining companies excessively with our experiment. In the CVs, the gender of our two applicants was independently randomized, with a 50% probability of being female or male. As a result, application pairs with either two females, two males, or with one female and one male were sent via e-mail to a specific employer. Our design thus required two profiles per gender and language region. We also independently randomized other features like the gender of the applicant's sibling and the gender of the teacher given as a reference person. In contrast, mother's occupation was randomized pairwise among the two applications per open position, implying that these applications had necessarily different values for mother's occupation. The latter was either homemaker or primary school teacher, each with a chance of 50%.

Father's occupation was also randomized pairwise (and independently of mother's occupation) and contained the following options: university professor (with 12.5% probability), an intermediate technical position (37.5%) matching the job type of the apprenticeship (e.g. mechanic), an intermediate commercial position (37.5%) matching the job type (e.g. sales manager), and an unskilled worker (12.5%). The idea was to consider high skilled, low skilled, as well as intermediate profiles, with the latter being related to the position to be filled. The skill level of intermediate profiles therefore varied depending on the tier and industry of the position. For instance, for a technical apprenticeship in the first, second, or third tier, father's intermediate technical occupation would either be a mechanic, a polymechanic, or an engineer. This implies substantial heterogeneity of educational achievements within the intermediate profiles for the sake of aligning father's occupation well with open apprenticeships. Some other CV features, such as motivational sentences and leisure activities, were also randomized pairwise in order to make sure that the same phrases and hobbies would not be used twice in applications sent to the same vacancy.

Employers advertise apprenticeship positions in specialized job portals, at least if they cannot be filled through professional or personal networks. In total, 3069 applications were sent out between August and mid October 2018 via e-mail to open positions posted on Switzerland's most popular online portal for apprenticeships.

During the data collection process, several issues arose. In August, we accidentally sent out applications to some positions that were from the previous year and thus not relevant for our fictitious candidates. In a few cases, the employers' e-mail addresses provided online contained typos or were not valid such that the applications could not be sent. In total, 129 observations were dropped due to such issues. Furthermore, while most employers received two applications as intended, 397 employers in Lausanne only received one application due to technical issues at the end of the application period (end of September until mid October). However, also in these cases, the application features were correctly randomized as described above.

A more serious concern was that five employers in the German speaking region detected that our applications were not related to existing students, having followed up on the candidates by consulting the schools. Even though these cases were excluded from the analysis, it cannot be ruled out that the information was communicated to other employers. If so, this could bias our results.

We envision several ways in which this may have occurred. One possibility is that employers started ignoring our candidates, thus biasing

any effects toward zero. Another possibility is that, once made aware of the experiment, employers decided to invite our candidates to an interview when they otherwise would not have done so. Such procedure would likewise reduce our ability to identify any empirical links between demographic attributes and acceptance rates. Nonetheless, as described below, our results for the whole experimental period present a very specific pattern whereby one particular parental occupation listed in the candidates profiles led to a great boost in the success rate of our female candidates. This pattern could hardly have been the result of intentional behavior. Robustness checks discussed below, where we restricted the sample to a time period when the detections were more likely to have had an effect in the way companies responded to our candidates, deliver a comparable response pattern as the one for the whole sample. We thus believe that these issues did not affect callback rates on a large scale.

A final incident (also in the German speaking part) concerns a situation where we accidentally sent out four applications to the same employer, resulting in the same applicant's name being sent twice. Even though no reaction by the employer was received, we immediately withdrew our applications when noticing the issue and excluded this employer from the sample, as well. All in all, we dropped 12 observations because of the issues mentioned. Our final evaluation data set thus consists of 2928 observations.

Most of the time, employers responded to our applications by e-mail, though phone calls were also frequent. We never answered the phone directly but regularly checked on the messages left by companies in the voicemail of our fictitious applicants' phone numbers. In 10 out of the comparably few instances when actual letters were sent as replies to our applications, they could not be delivered and were returned to employers, who then wrote e-mails to ask for a correct address. In these cases, we replaced the problematic addresses with new ones (which were then subsequently used for the continuation of the study). We apologized to the companies via e-mail and asked to have the letters sent to the new address or for the possibility to get the message via e-mail instead. These employers are kept in our evaluation sample, albeit excluding them leaves our results virtually unchanged.

If one of our applications received an invitation, which was either for a job interview, an assessment center, or a trial apprenticeship, we declined the offer within several days. In this case, the dependent variable, *employer response*, was coded as one, corresponding to a 'callback'. In the case of a negative response or no reaction on the part of the employer until to February 2019, the dependent variable was coded as zero.

Ethical questions and debriefing The methodology of correspondence testing raises ethical issues, as it necessarily involves the deception of recruiters assessing the electronic documents of our fictitious applicants. While ethical concerns are of first importance and have been addressed in the literature, see e.g. Riach and Rich (2004), it has also been recognized that carrying out research based on a correspondence testing methodology (or, more generally, on field experiments) requires breaking informed consent, see Blommaert, Coenders, and van Tubergen (2013). Indeed, informing participants a priori would invalidate the experiment.

Well-defined exceptions to informed consent have been established in law in a variety of countries (e.g. Sweden, see Bursell, 2007, and the USA, see Pager, 2007). In the discussion of ethical issues and correspondence testing, one argument often used in favor of the methodology is the relevance of the research question. Arguably, investigating the prevalence of discrimination is a pursuit worth following whose merits could outweigh the cost of not informing participants beforehand. Indeed, the use of deception has been defended on the grounds of the necessity to evaluate the effectiveness of anti-discriminatory legislation, see Banton (1997). Many courts, including e.g. the US Supreme Court, have endorsed 'tester' methodologies. (For legal practices, it is common to 'test' one company multiple times whereas the practice

of correspondence testing addresses many companies and tends to focus on particular employers only once.) Such practices have gained systematic support from US courts over time, see Pager (2007). In Sweden, initial rejections of the methodology from the Swedish Ethics Board were later overturned (thus aligning with many other OECD countries) after the use of testing results in legal proceedings against detected discriminatory practices, which demonstrated the usefulness and social relevance of the methodology, see Carlsson and Rooth (2012) for an example.

Even if an exception to the principle of informed consent is accepted, correspondence testing poses costs to employers, as recruiters spend time on evaluating fictitious candidates. However, if only a small number of applications is sent to each company and if invitations to interviews (or to a follow-up action) are swiftly declined, the time cost can be kept at a comparably small level, as argued in Wood, Hales, Purdon, Sejersen, and Hayllar (2009). In this study, we adhered to these practices, e.g. by sending out not more than two applications per employer.

While informing participants about an ongoing experiment would invalidate its results, ethical considerations may suggest informing participants *ex-post*. Debriefing practices nonetheless also have potential downsides, as discussed in Liebkind, Larja, Brylka, et al. (2016), Midtbøen (2014), and Pager (2007); for instance, they may invalidate future experiments. Zschirnt (2019) provides a thorough overview of how the discussion and practices surrounding correspondence testing have evolved in the literature. In our experiment, we debriefed companies once the data collection period was completed. In early March 2019, we sent e-mails with attached letters that explained the setup, purpose, and key findings of the experiment to employers that had received applications from our fictitious candidates. The vast majority of employers did not react to the debriefing. Among the 11 responses we received via e-mail, some expressed dissent and discontent with the fact they had been confronted with fictitious applications while others had critical comments or questions concerning the methodology, which we in turn answered in a subsequent e-mail. One reaction was positive and pointed out the importance of investigating discrimination.

5 Data and descriptive statistics

Our evaluation sample consists of 2928 observations and contains information about applications and (in anonymized form) employers. Application characteristics consist of apprenticeship tiers in terms of required qualifications, types in terms of gender orientation (female-dominated, gender neutral, male-dominated), applicant gender, parental occupation, the agglomeration in which the apprenticeship was located, and whether or not the fictitious applicant had moved from a different city (and thus had a certificate from a school located elsewhere). We also recorded the dates when an apprenticeship was posted (or, if unavailable, the date when it was found by the research team) and when an application was sent out.

Employer characteristics include categories for the (in many cases estimated) number of employees, the sector (i.e. public, trade and wholesale, manufacturing and goods, or services), the scale of the employer's operations (local, national, or international), the gender of the contact person in the company, whether or not there was an explicit anti-discrimination policy on the company's website, and the geographic distance (in kilometers) of the employer to the central station of the applicant's residential city. In addition to the characteristics, the data contain a binary outcome variable measuring employers' response to our applications and is one in case of an invitation to an interview, assessment center, or trial apprenticeship, and zero otherwise. The anonymized data set without the variable 'geographic distance' is publicly available at https://doi.org/10.7910/DVN/PIUJW4.

Table 1 provides the means of all characteristics but gender in the total sample, as well as separately by gender, which is the key

Table 1
Descriptive statistics by applicant gender.

	Total sample	female	male	t-test	
	mean	mean	mean	diff	p-val
employees: 1 to 20	0.48	0.48	0.47	0.00	0.88
employees: 21 to 50	0.24	0.24	0.24	−0.01	0.63
employees: 51 to 100	0.11	0.11	0.12	−0.00	0.74
employees: 101 to 250	0.09	0.09	0.08	0.01	0.52
employees: 251 to 500	0.03	0.03	0.03	−0.00	0.87
employees: 501 to 1000	0.02	0.02	0.02	−0.00	0.97
employees: more than 1000	0.04	0.04	0.04	0.00	0.67
sector: public	0.05	0.05	0.06	−0.00	0.69
sector: trade and wholesale	0.22	0.22	0.23	−0.01	0.48
sector: manufacturing and goods	0.13	0.12	0.13	−0.01	0.64
sector: services	0.60	0.61	0.59	0.02	0.27
distance to city center	16.37	16.22	16.55	−0.33	0.57
tier 1 job	0.35	0.34	0.36	−0.02	0.32
tier 2 job	0.36	0.38	0.35	0.03	0.15
tier 3 job	0.28	0.28	0.29	−0.01	0.63
type: gender-neutral	0.33	0.32	0.33	−0.01	0.56
type: female-dominated	0.25	0.27	0.22	0.05	0.00
type: male-dominated	0.43	0.41	0.45	−0.04	0.03
city: Bern	0.21	0.21	0.21	−0.00	0.91
city: Zurich	0.30	0.29	0.31	−0.02	0.30
city: Basel	0.11	0.13	0.09	0.03	0.00
city: Lausanne	0.38	0.37	0.39	−0.02	0.39
activity: regional	0.80	0.81	0.79	0.02	0.27
activity: national	0.12	0.11	0.13	−0.02	0.16
activity: international	0.08	0.08	0.08	0.00	0.95
antidiscrimination policy	0.21	0.22	0.20	0.02	0.23
contact: female	0.31	0.32	0.30	0.02	0.29
contact: male	0.33	0.32	0.34	−0.02	0.36
contact: unknown	0.36	0.36	0.36	−0.00	0.90
day job was published or found	29.08	29.00	29.17	−0.18	0.74
day of application	51.00	50.80	51.22	−0.42	0.50
father professor	0.13	0.12	0.14	−0.03	0.04
father intermediate	0.75	0.76	0.74	0.02	0.15
father unskilled worker	0.12	0.12	0.12	0.00	0.82
mother teacher	0.50	0.51	0.48	0.03	0.07
applicant has moved	0.49	0.50	0.48	0.02	0.29
number of observations	2928	1529	1399		

Note: Means of characteristics in the total, female, and male samples, as well as mean differences ('diff') between females and males and p-values of two sample t-tests ('p-val').

intervention variable of our experiment. It also contains mean differences across gender ('diff') and p-values ('p-val') of two sample t-tests. The characteristics' means are generally well balanced across gender as only few mean differences are statistically significant at the 5% level. We also test mean balance of all characteristics jointly based on the machine learning approach of Ludwig, Mullainathan, and Spiess (2017), which is outlined in Appendix B and provides no indication of imbalances, with a p-value of 98.4%.

Table 2 reports descriptives by parental occupation (rather than gender) as our second intervention variable of interest. In the first column, it displays the means of all characteristics but parental occupation for the group of applications with the mother being a teacher and the father being an unskilled worker ('mean'). Furthermore, it shows mean differences ('diff') between this reference group and other combinations of parental occupation, namely: mother is a teacher and father has an intermediate occupation (technical or commercial), mother is teacher and father is a university professor, mother is a homemaker and father is a low skilled worker, mother is a homemaker and father has an intermediate occupation (technical or commercial), and mother is homemaker and father is a university professor. P-values for the respective two sample t-tests are also reported ('p-val'). Again, the majority of mean differences is not statistically significant at the 5% level. We also apply the joint testing procedure of Ludwig et al. (2017) for the pairwise testing of mother is a teacher vs. mother is a homemaker, father has an intermediate occupation vs. father has a different occupation, and father is a professor vs. father is not a professor. The respective p-values are 5.2%, 91.6%, and 96.7%. By

Table 2
Descriptive statistics by parental occupation.

	mother teacher, father unskilled	mother teacher, father intermediate		mother teacher, father professor		mother homemaker, father unskilled		mother homemaker, father intermediate		mother homemaker, father professor	
	mean	diff	p-val	diff	p-val	diff	p-val	diff	p-val	diff	p-val
employees: 1 to 20	0.56	−0.09	0.03	−0.17	0.00	−0.10	0.07	−0.08	0.06	−0.13	0.02
employees: 21 to 50	0.22	0.03	0.46	0.01	0.89	0.02	0.61	0.0	0.74	0.03	0.47
employees: 51 to 100	0.09	0.02	0.30	0.07	0.04	0.05	0.16	0.03	0.20	−0.01	0.74
employees: 101 to 250	0.06	0.03	0.13	0.05	0.08	0.00	0.98	0.03	0.13	0.08	0.01
employees: 251 to 500	0.04	−0.02	0.16	0.00	0.91	0.00	0.90	−0.02	0.28	−0.03	0.13
employees: 501 to 1000	0.01	0.01	0.14	0.03	0.06	0.01	0.23	0.01	0.27	0.02	0.13
employees: more than 1000	0.02	0.02	0.12	0.00	0.83	0.01	0.73	0.02	0.21	0.03	0.13
sector: public	0.02	0.04	0.00	0.03	0.10	0.07	0.00	0.03	0.01	0.04	0.03
sector: trade & wholesale	0.21	0.01	0.78	−0.01	0.83	0.02	0.65	0.03	0.44	−0.03	0.54
sector: manufacturing & goods	0.16	−0.03	0.29	−0.06	0.12	−0.05	0.19	−0.03	0.28	−0.02	0.65
sector: services	0.61	−0.01	0.74	0.03	0.52	−0.04	0.44	−0.02	0.56	0.00	0.99
distance to city center	18.96	−2.72	0.04	−3.29	0.05	−2.28	0.18	−3.03	0.02	−1.22	0.46
tier 1 job	0.36	−0.02	0.61	0.05	0.38	−0.01	0.82	−0.01	0.76	0.02	0.64
tier 2 job	0.37	−0.01	0.88	−0.03	0.60	0.00	0.98	−0.01	0.77	−0.04	0.38
tier 3 job	0.26	0.03	0.48	−0.02	0.68	0.01	0.83	0.02	0.52	0.02	0.67
type: gender-neutral	0.30	0.02	0.63	0.05	0.31	0.01	0.77	0.04	0.29	0.03	0.49
type: female-dominated	0.25	−0.01	0.84	−0.00	0.97	0.00	0.96	−0.01	0.83	−0.01	0.86
type: male-dominated	0.45	−0.01	0.80	−0.05	0.35	−0.02	0.76	−0.03	0.43	−0.03	0.62
city: Bern	0.21	0.00	0.90	0.00	0.97	−0.00	0.99	−0.01	0.88	0.04	0.43
city: Zurich	0.31	−0.02	0.66	0.01	0.92	−0.02	0.71	−0.01	0.73	−0.01	0.79
city: Basel	0.10	0.00	0.99	0.02	0.55	0.03	0.42	0.01	0.84	0.00	0.94
city: Lausanne	0.37	0.01	0.76	−0.03	0.60	−0.01	0.86	0.01	0.75	−0.02	0.64
activity: regional	0.83	−0.03	0.36	−0.04	0.36	−0.03	0.50	−0.04	0.26	−0.05	0.25
activity: national	0.11	0.00	0.88	0.03	0.38	0.01	0.85	0.01	0.62	0.01	0.74
activity: international	0.06	0.03	0.20	0.01	0.78	0.02	0.42	0.02	0.25	0.04	0.19
antidiscrimination policy	0.17	0.04	0.18	0.06	0.16	0.01	0.88	0.04	0.22	0.07	0.09
contact: female	0.26	0.05	0.16	0.05	0.27	0.04	0.46	0.05	0.19	0.05	0.29
contact: male	0.34	−0.01	0.76	−0.02	0.70	0.02	0.74	−0.02	0.65	−0.00	0.95
contact: unknown	0.39	−0.04	0.32	−0.03	0.51	−0.05	0.31	−0.03	0.45	−0.05	0.36
day job was published/found	28.44	0.66	0.60	0.82	0.62	0.18	0.91	1.04	0.41	−0.79	0.61
day of application	49.45	1.98	0.17	−0.48	0.80	0.80	0.66	2.38	0.10	−1.66	0.37
applicant: female	0.55	0.00	0.98	−0.06	0.25	−0.03	0.53	−0.03	0.42	−0.08	0.11
applicant has moved	0.48	0.01	0.76	−0.01	0.90	0.02	0.72	0.02	0.66	−0.02	0.68
number of observations	163	1119		176		197		1076		197	

Note: 'mother teacher, father unskilled' provides the means of characteristics in the reference group (mother teacher, father unskilled worker), the other columns provide the mean differences ('diff') compared to the baseline group and the p-values ('p-val'), respectively.

and large, characteristics thus appear satisfactorily balanced across our intervention variables of interest, namely applicant gender and parental occupation. For the single variable of mother's occupation, however, balance is almost rejected at the 5% level of significance, but this *p*-value does not account for the fact that we run the Ludwig et al. (2017) test for multiple hypotheses. In any case, our empirical results presented in Section 6 are very similar when conditioning or not conditioning on application and employer characteristics to control for observed imbalances.

6 Results

Our paper aims to find out whether gender and parental occupation have an effect on callback rates for the young applicants of the Swiss apprenticeship market. To this purpose, we run versions of the following regression:

$$\text{callback}_i = \alpha + \beta \, \text{gender}_i + \theta_0 \, \text{mother's\_occupation}_i$$
$$+ \theta_1 \, \text{mother's\_occupation}_i \times \text{gender}_i +$$
$$\delta_0 \, \text{Father's\_Occupation} + \delta_1 \, \text{father's\_occupation} \times \text{gender}_i$$
$$+ X_i \, \phi + u_i$$

where callback_i is a dummy variable for the callback outcome for person i (taking the value one when this person received an invitation to an interview, assessment center, or trial apprenticeship, and zero otherwise), gender_i is the applicant's gender (with 0 for males and

1 for females); additional terms control for the mother and father's occupations, with further interactions of those with gender. Finally, we include a vector of control variables for the company to which the application was submitted (such as sector, distance to city center, number of employees, and scope of operations — national or international), summarized in X_i. These controls are listed in Tables 1 and 2.

Since – for each company – gender and parental occupations were randomized, in order to focus on one particular effect we will run regressions where only that effect is controlled for. Take gender, for example. To find out if callbacks favor boys versus girls, we begin by running a regression that includes only that control. Alternatively, if wanting to find out the impact of maternal occupation, we will run regressions where only gender and maternal occupation (and their interaction) are controlled for. Since the other controls were randomized, these procedures deliver unbiased estimates of the relevant parameters.

We begin our analysis by examining whether callback rates differ across applicant gender. Table 3 shows our results for gender effects both with and without including control variables. Looking at the estimates ('est') reported in the first column of the left panel, the mean callback rate for male applicants amounts to 27% (29% when including control variables). Callback rates vary quite broadly in the literature, depending on the specific candidate type and occupation considered. In Riach and Rich (2006) for example, callback rates for chartered accountants varied between 6.5 and 9.7% whereas those of computer analysts ranged from 13.8 to 23.1% (see Table 2 in that paper). For most of the occupations and candidate types considered in

Table 3
Gender differences in callbacks.

	est	boot se	pval	est	boot se	pval
Gender effect	No controls			With controls		
male (mean/intercept)	0.272			0.294		
female (diff)	0.034	0.018	0.050	0.032	0.017	0.055

Table 4
Effects of gender and father's occupation.

	est	boot se	raw pval	adj pval
female: father unskilled worker (mean)	0.242	0.031	0.000	
female: father intermediate	0.060	0.033	0.070	0.331
female: father professor	0.161	0.047	0.001	0.003
male: father unskilled worker	0.028	0.046	0.533	0.777
male: father intermediate	0.032	0.035	0.354	0.777
male: father professor	0.017	0.043	0.700	0.777

Note: estimates of (differences in) callback rates for the total sample, without control variables. 'est' provides the callback rate for the group 'female: father unskilled worker', as well as the differences in callback rates of all other groups relative to 'female: father unskilled worker'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Table 5
Effects of gender and father's occupation with controls.

	est	boot se	raw pval	adj pval
female: father unskilled worker (intercept)	0.373	0.088	0.000	
female: father intermediate	0.061	0.032	0.054	0.903
female: father professor	0.152	0.044	0.001	0.083
male: father unskilled worker	0.040	0.045	0.378	0.981
male: father intermediate	0.026	0.032	0.416	0.999
male: father professor	0.003	0.043	0.941	1.000

Note: estimates of (differences in) callback rates for the total sample, with control variables. 'est' provides the callback rate for the group 'female: father unskilled worker', as well as the differences in callback rates of all other groups relative to 'female: father unskilled worker'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Table 6
Effects of gender and mother's occupation.

	est	boot se	raw pval	adj pval
female: mother teacher (mean)	0.317	0.016	0.000	
female: mother home	−0.022	0.020	0.281	0.292
male: mother teacher	−0.040	0.023	0.085	0.121
male: mother home	−0.050	0.019	0.009	0.051

Note: estimates of (differences in) callback rates for the total sample, without control variables. 'est' provides the callback rate for the group 'female: mother teacher', as well as the differences in callback rates of all other groups relative to 'female: mother teacher'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Petit (2007), callback rates went from 30 to 70%, a very wide range (their Table 3). Becker S. O. et al. (2019) is a natural reference for us and callback rates there were in the range 15%–17% for Switzerland (Table 1). In comparison with the latter and the literature in general, we believe that our candidates were quite successful.

With a p-value of exactly 5% (based on cluster bootstrapping), the differential callback rate for female applicants amounts to 3.4 pp. Including controls into our estimations, we find this differential callback rate to be slightly lower, 3.2 pp, with a p-value of 5.5%. As we will see below, the higher callback rate for girls detected in Table 3 is in fact driven by a particular type of parental occupation, when the father is a university professor. We proceed to study the impact of paternal and maternal occupations next and will later come back to this point.

As a next step, we analyze whether the gender effect on callback rates differs by parental occupation. We begin by interacting gender with father's occupation. The reference category is females with an unskilled worker as a father. The average employer response (i.e. the share of invitations) for the reference category is reported ('est') and amounts to roughly 24%. For the other 5 categories defined by combinations of gender and paternal occupation, we report the respective difference to the reference category ('est'), along with bootstrap standard errors ('boot se') and conventional p-values ('raw p-val') based on a t-test. However, these p-values do not take into account multiple hypothesis testing, i.e. the fact that we simultaneously test five differences. This is problematic because the likelihood of spuriously rejecting one or even several null hypotheses generally increases in the number of hypotheses tested. We therefore adjust the p-values of each difference for multiple testing ('adj p-val') using the stepdown approach of Romano and Wolf (2005, 2016). The latter exploits the coefficient estimates in the bootstrap samples in order to compute test statistics that are related to the maximum statistical significance among all coefficients, which in turn permits adjusting the p-values of individual coefficients.

As shown in Table 4, we find that having a professor as a father raises callback rates for female applicants by about 16 pp. The effect is statistically significant when considering both conventional p-values ('raw pval') as well as adjusted p-values ('adj pval'). When controlling for covariates in the regression of callbacks on the interactions between gender and father's occupation, our findings remain robust, see Table 5. The reference category – females whose father is an unskilled worker – has an average callback rate of 37.3%. While almost all interactions of gender and father's occupation are statistically insignificant when accounting for multiple hypothesis testing, the coefficient on female

applications with a professor as father amounts to 15.2 pp and is statistically significant at the 10% level.

We then proceed to examine the gender differential according to the maternal occupation of applicants. For this case, our reference category are female applicants with a mother who is a teacher, whose callback rate amounts to 31.7%. Again, we bootstrap standard errors and compute conventional p-values as well adjusted p-values for the remaining three categories of gender and parental occupation. As the results in Table 6 show, we generally do not find that callback rates of males or females differ substantially across the occupation of their mother. As an exception, we note that the negative coefficient on applications of males whose mother is a homemaker is almost statistically significant at the 5% even when considering the adjusted p-values. Yet, this effect is in absolute terms much smaller than the coefficient on female applications with the father being a professor in Tables 4 and 5. Running the regression conditional on covariates (Table 7) does not importantly change the coefficient sizes, but any effect (including that for male applications whose mother is a homemaker) are now far from being statistically significant. Therefore, our findings do all in all not go against the hypothesis that employers do on average not distinguish between male and female applicants for the empirically most relevant case that paternal occupation is not a professor.[11]

It must be noted that, even though being the daughter or son of a professor bears comparably little empirical relevance, our finding points to distinct signaling effects for females and males, at least in this specific case. Our results therefore provide some support for the

---

[11] The absence of a significant gender effect on callback rates (excluding professors) is robust to differences in the variance of unobserved determinants of productivity across genders, see the discussion in Heckman (1998) and Heckman and Siegelman (1993). When applying the methodology of Neumark (2012) to decompose the total gender effect into its level and variance components, we find that the level effect, i.e. the component associated with (taste-based or statistical) discrimination, is very close to zero and statistically insignificant. The variance component is not statistically significant either. Results are provided in Table 20 in Appendix D.

Table 7
Treatment effects of gender and mother's occupation with controls.

	est	boot se	raw pval	adj pval
female: mother teacher (intercept)	0.447	0.086	0.000	
female: mother home	−0.027	0.019	0.154	0.989
male: mother teacher	−0.049	0.024	0.038	0.852
male: mother home	−0.055	0.019	0.005	0.825

Note: estimates of (differences in) callback rates for the total sample, with control variables. 'est' provides the callback rate for the group 'female: mother teacher', as well as the differences in callback rates of all other groups relative to 'female: mother teacher'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

enforcement of blind applications that do not reveal personal attributes like parental occupation in order prevent differential treatment due to signaling.

Our analysis thus far shows that daughters of university professors benefit from a quantitatively substantial increase in callbacks. We now revisit the results on gender and callbacks from Table 3. We consider the same empirical specification but now divide the sample into two subsets: one without university professors as fathers and the other including only including professors. The results are provided in Table 8. As the upper panel indicates, for the subsample devoid of parents who are university professors, the estimate of the female coefficient is now smaller and statistically insignificant: either 2 pp when no additional controls are used, or 2.4 pp with controls. The bottom panel for the subset with professors as parental occupation yields a completely different picture. The coefficient estimate amounts to 12.5 pp (with or without additional controls) and is statistically significant at the 5% level even when adjusting for multiple hypothesis testing.

Table 8 suggests comparable callback rates of girls and boys except when their father is a professor. Nonetheless, this equality in the average callback rates could mask a reality whereby each gender would do disproportionately well in occupations with a matching pattern in terms of gender dominance. This would correspond to the findings typified earlier characterizing the experimental results for adult persons and gender-matched applications. To check for this, we examine whether callback rates by gender remain statistically identical across the different occupation types in our sample: male- or female-dominated, or gender-neutral. Results are presented in Table 9. We present results separately for the subsample devoid of applicants whose father is a professor (top panel) and for the subsample exclusively composed of those cases (bottom panel).

Table 9 shows that the result of gender neutrality for average callback rates remains even when we consider the gender-type of occupations targeted. Indeed, for the top panel of the table, the differential callback rate for girls (compared to boys) is at most 3.4 pp and never statistically significant at conventional levels. From the results applying to adult persons, we would have expected girls to have an advantage over boys in female-dominated professions and to experience a disadvantage in male-dominated ones. And, further, for these two effects to be quantitatively substantive. This is not what we find.

These results above do not extend to the subsample of applicants whose father is a professor. There (bottom panel), once again, parental profession benefits girls in a very large and statistically significant way relative to boys in female-dominated occupations. The effect is smaller and no longer statistically significant for neutral occupations and even closer to zero (and statistically insignificant) in male-dominated occupations.

The absence of important gender differences in callbacks considering parental background (safe for the noted exception) suggests that employers' decisions at the apprenticeship level do not contribute to the reinforcement of the intergenerational persistence of income. That would be the case if employers favored applicants with more educated parents and did particularly so for boys. Our results clearly show the

opposite: in the empirically seldom case of having a university professor as a father, it is girls that are disproportionately favored.

In order to assess the robustness of our results, we next perform a power analysis. We focus first on the results on gender specifically in relation to Tables 8 and 9. In Graph 1, we plot power against the size of the female differential effect for different sample sizes. The left panel presents results for a two-sided equivalence test in relation to the difference in callback rates of girls versus boys – as in Table 8. The right panel portrays the results of one-sided tests of whether e.g. girls would do better than boys in female dominated professions (or the opposite in male dominated occupations) – as in Table 9 which, however, provides the p-values of two-sided tests.

As the left panel of Graph 1 shows, for the sample devoid of professors as fathers we have 80% power to detect effects just under 5 pp. This is a large threshold relative to the effects found in our data (of only 2 pp). For the much smaller sample of parents who are professors, with only 373 observations, we would only have 80% power for effects of about 15 pp. Yet, it is in the smaller sample that we detect very large and statistically significant effects. The fact that we consistently find large and statistically significant effects in the sample with professor as fathers suggests that the low and statistically insignificant differential callback rate otherwise experienced by girls relative to boys is not the consequence of low power. Further, and although correspondence tests and vignette studies differ, our results align with the findings of Fossati et al. (2020) for the same labor market.

As the literature review indicated, former correspondence tests do not provide a clear prediction for whether girls or boys would receive higher callback rates on average across occupations. What the literature clearly points to is ingrained gender stereotyping in employer behavior, with women doing better than men in female dominated occupations (and vice-versa). We therefore turn to Table 9 and the right panel of Graph 1. Restricting attention to the subsample without professors, with 80% power we would detect effects of 6.6 pp and 9.6 pp for the male-dominated and female-dominated occupations, respectively. The differential callback rate in female-dominated occupations reported in the related correspondence study for the Swiss labor market of Becker S. O. et al. (2019) was roughly 8 pp — which is of similar magnitude as the thresholds above. Therefore, our findings that girls and boys do similarly well in terms of callback rates even when considering the gender type of occupations is clearly not driven by low power.

Graph 2 plots similar power estimates but now looking at parental professions. The left panel corresponds to paternal occupations and the right one illustrates maternal ones. The corresponding Tables are Tables 4 and 6, respectively. Overall, our data allows us to detect economically significant effects (e.g. 8 pp or more, looking at maternal occupations, and roughly 10 to 12 pp for paternal occupations) with 80% power. The magnitude of the coefficients presented in Tables 4 and 6 is generally lower than these thresholds except for paternal occupation as a professor. Once again, the latter are the estimates where statistical power would be the lowest, by construction. The experimental literature offers less guidance here due to the scarcity of work on the apprenticeship labor market (where mentioning parental occupation is natural). Nonetheless, we interpret the aforementioned pattern of results as indicating that, overall, Swiss employers posting apprenticeship openings are making their hiring decisions mostly based on applicant merit and devoid of gender stereotyping and parental background considerations.

In Section 4, we discussed that to the best of our knowledge five employers detected that our applications were not related to existing students. Four detections were related to applications sent out between August 28th and September 7th, only one detection to applications in October. As a robustness check, we therefore run our main analysis for the month of September only, to investigate whether a potential communication among employers about the detection of fictitious applications affected our main findings. Even though we cannot rule out that some employers exchanged information on this issue and adapted
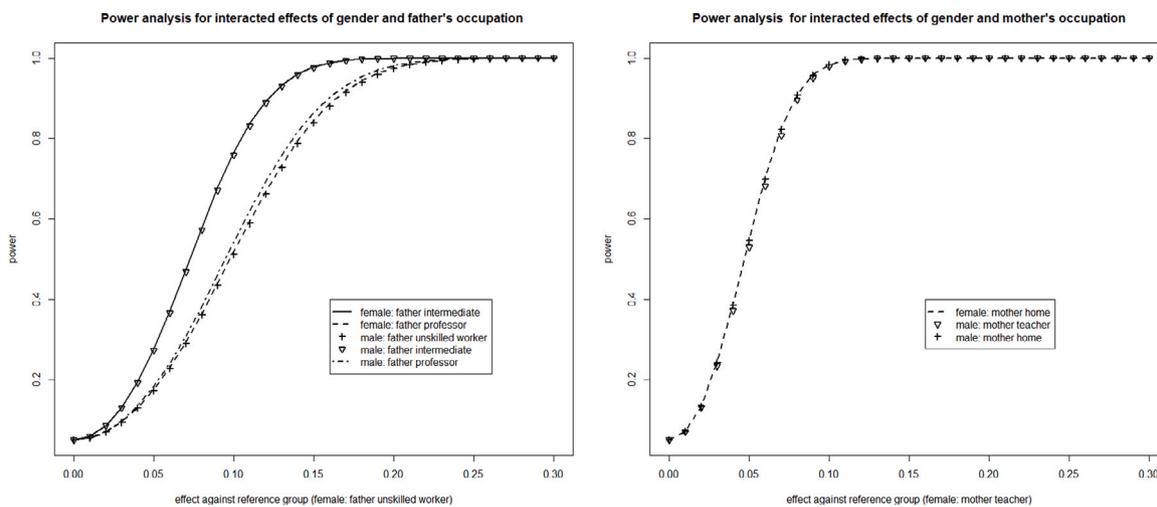
Table 8
Gender differences in callbacks across sub-samples.

	est	boot se	raw pval	adj pval	est	boot se	raw pval	adj pval
Father is not professor	No controls				With controls			
male (mean/intercept)	0.274				0.397			
female (diff)	0.020	0.018	0.274	0.675	0.024	0.017	0.162	0.534
number of observations	2555				2555			
Father is professor	No controls				With controls			
male (mean/intercept)	0.259				0.373			
female (diff)	0.125	0.055	0.023	0.022	0.125	0.051	0.014	0.016
number of observations	373				373			

Note: estimates of (differences in) callback rates across gender when father is reported or not reported to be a professor, without and with control variables. 'est' provides the callback rate or intercept for males and the difference in call back rates between females and males. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing of differences by gender, professor, and the gender–professor-interaction.
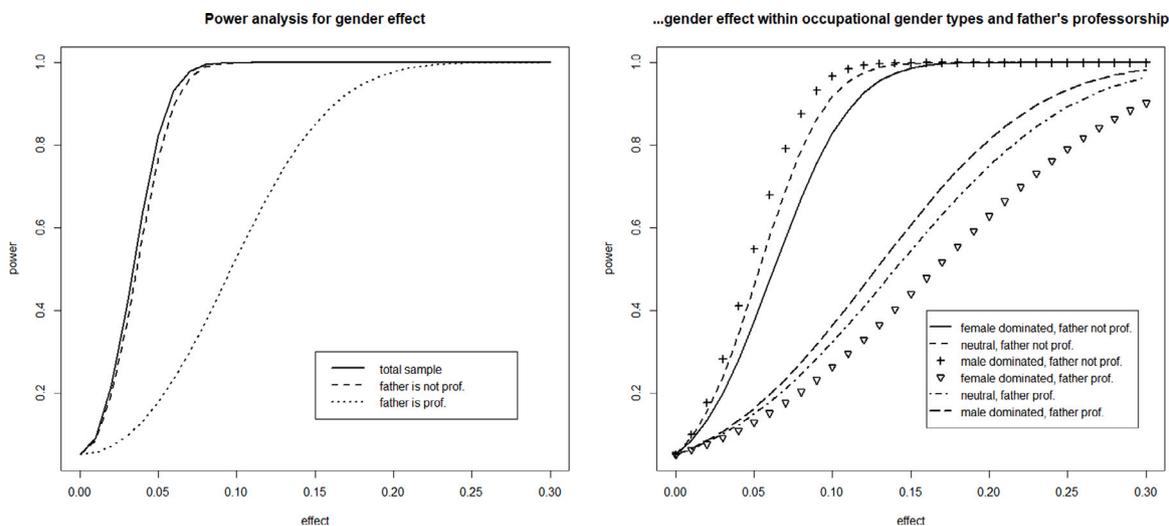
Graph 1. Power analysis for gender effect.

Graph 2. Power analysis for parental occupation and gender.

Table 9

Gender differences in callbacks by professor status and occupational gender types.

	est	boot se	raw pval	adj pval	est	boot se	raw pval	adj pval	est	boot se	raw pval	adj pval
Father is not professor	Female-dominated				Neutral				Male-dominated			
male (mean/intercept)	0.191				0.283				0.307			
female (diff)	0.034	0.033	0.302	0.308	0.030	0.031	0.329	0.732	0.016	0.029	0.589	0.707
number of observations	626				836				1093			
Father is professor	Female-dominated				Neutral				Male-dominated			
male (mean/intercept)	0.089				0.261				0.349			
female (diff)	0.260	0.088	0.003	0.004	0.116	0.091	0.206	0.221	0.049	0.082	0.546	0.707
number of observations	92				128				153			

Note: estimates of (differences in) callback rates across gender when father is reported or not reported to be a professor by type, without control variables. 'est' provides the callback rate or intercept for males and the difference in call back rates between females and males. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing of differences by gender, professor, and the gender–professor-interaction.

their response behavior accordingly, our results suggest that this was not a widespread phenomenon. As can be seen by comparing Tables 4 and 6 and Table 15 in Appendix C, the results are qualitatively in line with those of the complete sample. This is visible both in the broad similarity of point estimates of company responses to our different applicant types as well as in the fact that estimates of the female–professor-interaction effect remain also quantitatively not too different from those in the main sample, albeit now less precisely estimated.

In a next step, we investigate the heterogeneity of our results across specific characteristics, starting with language regions. To this end, we analyze the effects of the various combinations of applicants' gender and parental education (as considered in Tables 4 and 6) separately for the German (Basel, Bern, and Zurich) and French (Lausanne) speaking regions, see Table 11 in Appendix C. Having a university professor as father has a positive impact on the callback rate of female applicants in either language group, but this effect is on average larger in the German speaking sample. However, it cannot be rejected at conventional levels of significance that the respective estimates in the French and German speaking samples are actually the same, in particular when accounting for further multiple hypothesis testing issues introduced by splitting by language region. Any other difference in callback rates relative to the reference group is insignificant in either language group.

We consider three further heterogeneity checks based on conducting the main analysis separately by tiers (related to levels of qualifications), apprenticeship types in terms of gender orientation, or or company size (number of employees), see Tables 12, 13, and 14 in Appendix C. It appears that the female–professor interaction effect found in the main sample is predominantly driven by the lower (first and second) tiers, female-dominated apprenticeships, and smaller firms with up to 50 employees. However, we abstain from making strong claims about differences across subgroups, due to issues of multiple hypothesis testing.

The patterns of effect heterogeneity are by and large confirmed when investigating callbacks across applicant gender in subsamples with and without professor as paternal occupation similar to Table 8, however, separately by language regions, tiers, types, or company size. As can be seen from Tables 17, 18, and 19 in Appendix C, and Table 9, no statistically significantly different callback rates across females and males occur in the subsamples excluding professorship. In the subsample with professorship, the callback rate of females is more than 17 pp higher than that of males in the German speaking regions, while the difference is much closer to zero and statistically insignificant in the French speaking region. Furthermore, the female premium is 15 pp among tier one and two apprenticeships, but virtually nonexistent in the third tier. The gender difference is more pronounced among smaller companies with up to 50 employees. Finally, among female-dominated types, the respective effect amounts to 26 pp, but shrinks in magnitude and significance when going to gender neutral and male-dominated occupations.

We conclude our results section by pointing out that the striking and gendered difference in callback rates for candidates whose father is a university professor does not reflect a reaction to a hypothetical situation whereby girls with such a parental background applied seldomly to apprenticeship positions compared to boys. Indeed, as the picture in Appendix F shows, a university professor (presumably someone with a Doctoral degree) is a rare occurrence for *both* girls and boys. Using data from the European Social Survey for individuals who followed an apprenticeship and are over the age of 25 (thus presumably having completed their education), comprising the 2010, 2012, 2014, and 2016 waves, we were able to get information on the professions of the apprentices' fathers. As shown in Appendix F, the share of fathers with a doctoral degree is 0.80% for girls and 0.58% for boys. Parental doctoral education, if anything, is more common for girls than boys.

7 Conclusion

We investigated the effect of gender interacted with parental occupation on callback rates for applications to apprenticeship positions by means of a correspondence test. Sending out approximately 3000 fictitious applications in four regions of Switzerland, our intervention variables did not affect callbacks in a statistically significant way in most cases. We therefore found no robust evidence of employers applying differential treatment to applicants w.r.t. to gender or parental occupation in the Swiss apprenticeship market. The one exception was when the applicant stated having a university professor as father, which boosted callbacks for females in a statistically significant way, even when accounting for multiple hypothesis testing, but not for males. Albeit paternal professorship is an empirically rare case, this finding points to the possibility of signaling effects of parental occupation among female applications. This suggests that applications should ideally be blind and not reveal socio-economic information in order to maximize fairness.

As outlined in the introduction, gender occupational segregation is often the object of policy focus as it is perceived to be a potential source of gender inequality in labor market outcomes. Our results represent rather positive news for the Swiss apprenticeship market: companies do not appear to contribute to an early onset of gender occupational segregation through gender-biased callbacks — at least not to a level that we can statistically detect. To put this finding into perspective, our power analysis suggests that we can detect a gender effect on the call back rate which is as small as 5 percentage points with a probability slightly higher than 80% in our total sample. Therefore, gender occupational segregation at the apprenticeship level seems to be predominantly driven by supply-side effects. Consequently, policies aimed at fostering gender equality across occupations should focus on removing gender related educational or cultural barriers influencing occupational choices at young ages.

Table 10
List of occupations.

id	Occupations in German	Occupations in French	Tendency
1	Bäcker/in-Konditor/in-Confiseur/in EFZ	Boulanger/ère-pâtisier/ière-confiseur/euse CFC	N
2	Coiffeur/-euse EFZ Hairdresser	Coiffeur/euse CFC	F
3	Detailhandelsassistent/in EBA Retail assistant	Assistant/e du commerce de détail AFP	N
4	Fachmann/-fachfrau Betriebsunterhalt EFZ	Agent/e d'exploitation CFC	M
5	Gärtner/in EFZ	Horticulteur/trice CFC	M
6	Koch/Köchin EFZ	Cuisinier/ière CFC	N
7	Logistiker/in EFZ	Logisticien/ne CFC	M
8	Restaurationsfachmann/-frau EFZ	Spécialiste en restauration CFC	N
9	Sanitärinstallateur/in EFZ	Installateur/trice sanitaire CFC	M
10	Schreiner/in EFZ	Charpentier/ière CFC	M
11	Montage-Elektriker/in EFZ	Electricien/ne de montage CFC	M
12	Automobil-Fachmann/-frau EFZ Automotive professionals	Mécanicien/ne en maintenance d'automobiles CFC	M
13	Maurer/in EFZ	Maçon/ne CFC	M
14	Polymechaniker/in EFZ, G-Profil	Polymechanicien/ne Profil G CFC	M
15	Dentalassistent/in EFZ	Assistant/e dentaire CFC	F
16	Fachmann/-fachfrau Betreuung EFZ	Assistant/e socio-éducatif/ve CFC	F
17	Kaufmann/-frau EFZ, B-Profil	Employé/e de commerce CFC, formation de base	N
18	Medizinische/r Praxisassistent/in EFZ	Assistant/e médical/e CFC	F
19	Pharma-Assistent/in EFZ	Assistant/e en pharmacie CFC	F
20	Detailhandelsfachmann/-frau EFZ, Beratung Retail professional	Gestionnaire du commerce de détail CFC, Conseil à la clientèle	N
21	Fachmann/-fachfrau Gesundheit EFZ	Assistant/e en soins et santé communautaire CFC	F
22	Automobil-Mechatroniker/in EFZ	Mécatronicien/ne d'automobiles CFC	M
23	Elektroinstallateur/in EFZ	Installateur/trice-électricien/ne CFC	M
24	Zeichner/in EFZ	Dessinateur/trice CFC	M
25	Metallbauer/in EFZ	Constructeur/trice métallique CFC	M
26	Kaufmann/-frau EFZ, E-Profil	Employé/e de commerce CFC, formation élargie	N
27	Informatiker/in in Applikationsentwicklung EFZ	Informaticien/ne en développement d'applications CFC	M
28	Informatiker/in in Betriebsinformatik EFZ	Informaticien/ne en informatique d'entreprise CFC	M
29	Informatiker/in in Systemtechnik EFZ	Informaticien/ne en technique des systèmes CFC	M
30	Polymechaniker/in EFZ, E-Profil	Polymécanicien/ne Profil E CFC	M

Note: Gender Tendency: N = Neutral, M = Male, F = Female. Occupational tiers: Tier 1: id: 1–12; Tier 2: ids 13–22; Tier 3: ids 23–30.

Appendix A. List of occupations

See Table 10.

Appendix B. Machine learning-based balance test

For jointly testing mean balance of all characteristics across the intervention variables gender or parental occupation as discussed in Section 5, we apply the machine learning-based test suggested by Ludwig et al. (2017). It is based on the intuition that the problem of obtaining too many significant results when testing multiple hypotheses (e.g. mean differences in multiple characteristics across gender), or false positives, is similar to the concern of overfitting in machine learning.

Applying the machine learning logic, we split our sample into training and testing data. In the training data, we run a lasso logit regression of the respective intervention variable on the characteristics using the 'rlogit' command with its default values in the 'hdm' package by Chernozhukov, Hansen, and Spindler (2015) for the statistical software 'R'. We then use the obtained coefficients for predicting the intervention in the test data and compare the prediction to the actual intervention variable to compute the mean squared error (MSE). We use 5-fold cross-validation, such that the roles of training and test data are swapped, and take the average of the 5 MSEs obtained. In a next step, we randomly relabel (or permute) the actual intervention and re-estimate the MSE using the same procedure. Repeating the permutation 999 times, we compute the p-value for the joint significance of the characteristics as the share of permutation based MSEs that are lower than the MSE with the correct coding of the intervention. The permutation test's intuition is that, if the characteristics are balanced across the intervention, relabeling does not systematically affect (i.e. increase) the MSE. If, on the other hand, characteristics are predictive for the intervention, the correct coding of the latter should likely entail a smaller MSE than the permuted versions.

Appendix C. Additional tables

See Tables 11–19.

Appendix D. Neumark (2012) correction for unobservable determinants of productivity

See Table 20.

Table 11

Treatment effects of gender and parental occupation by language regions.

	est	boot se	raw pval	adj pval
German language region				
female: mother teacher, father unskilled worker (mean)	0.232	0.056	0.000	
female: mother teacher, father intermediate	0.112	0.062	0.069	0.454
female: mother teacher, father professor	0.222	0.089	0.012	0.032
female: mother home, father unskilled worker	0.127	0.081	0.116	0.385
female: mother home, father intermediate	0.096	0.058	0.097	0.596
female: mother home, father professor	0.259	0.083	0.002	0.011
male: mother teacher, father unskilled worker	0.072	0.087	0.408	0.802
male: mother teacher, father intermediate	0.118	0.063	0.061	0.405
male: mother teacher, father professor	0.068	0.080	0.399	0.802
male: mother home, father unskilled worker	0.030	0.080	0.707	0.853
male: mother home, father intermediate	0.064	0.060	0.291	0.802
male: mother home, father professor	0.035	0.078	0.650	0.853
number of observations	1815			
French language region				
female: mother teacher, father unskilled worker (mean)	0.121	0.059	0.038	
female: mother teacher, father intermediate	0.161	0.066	0.014	0.283
female: mother teacher, father professor	0.212	0.102	0.038	0.146
female: mother home, father unskilled worker	0.041	0.084	0.626	0.871
female: mother home, father intermediate	0.081	0.064	0.208	0.737
female: mother home, father professor	0.114	0.094	0.224	0.555
male: mother teacher, father unskilled worker	0.057	0.095	0.545	0.871
male: mother teacher, father intermediate	0.069	0.064	0.279	0.811
male: mother teacher, father professor	0.040	0.090	0.657	0.871
male: mother home, father unskilled worker	0.193	0.101	0.057	0.189
male: mother home, father intermediate	0.094	0.066	0.151	0.633
male: mother home, father professor	0.136	0.093	0.146	0.453
number of observations	1113			

Note: estimates of (differences in) callback rates per language region, without control variables. 'est' provides the callback rate for the group 'female: mother teacher, father unskilled worker', as well as the differences in callback rates of all other groups relative to 'female: mother teacher, father unskilled worker'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Table 12

Treatment effects of gender and parental occupation by tiers.

	est	boot se	raw pval	adj pval
Tiers 1 and 2				
female: mother teacher, father unskilled worker (mean)	0.131	0.044	0.003	
female: mother teacher, father intermediate	0.152	0.049	0.002	0.116
female: mother teacher, father professor	0.278	0.076	0.000	0.002
female: mother home, father unskilled worker	0.108	0.068	0.111	0.287
female: mother home, father intermediate	0.108	0.047	0.021	0.287
female: mother home, father professor	0.275	0.073	0.000	0.003
male: mother teacher, father unskilled worker	0.089	0.072	0.215	0.328
male: mother teacher, father intermediate	0.116	0.049	0.017	0.236
male: mother teacher, father professor	0.123	0.066	0.064	0.223
male: mother home, father unskilled worker	0.133	0.072	0.064	0.211
male: mother home, father intermediate	0.086	0.047	0.069	0.328
male: mother home, father professor	0.090	0.067	0.178	0.328
number of observations	2097			
Tier 3				
female: mother teacher, father unskilled worker (mean)	0.321	0.090	0.000	
female: mother teacher, father intermediate	0.102	0.099	0.300	0.967
female: mother teacher, father professor	0.100	0.147	0.497	0.972
female: mother home, father unskilled worker	0.079	0.130	0.545	0.986
female: mother home, father intermediate	0.062	0.094	0.510	0.986
female: mother home, father professor	0.049	0.123	0.691	0.986
male: mother teacher, father unskilled worker	0.079	0.157	0.616	0.986
male: mother teacher, father intermediate	0.046	0.098	0.635	0.986
male: mother teacher, father professor	−0.071	0.126	0.571	0.986
male: mother home, father unskilled worker	0.012	0.134	0.929	0.986
male: mother home, father intermediate	0.057	0.099	0.567	0.986

Table 12 (*continued*).

	est	boot se	raw pval	adj pval
male: mother home, father professor	0.058	0.128	0.651	0.986
number of observations	831			

Note: estimates of (differences in) callback rates for tier 1 and 2 vs. tier 3, without control variables. 'est' provides the callback rate for the group 'female: mother teacher, father unskilled worker', as well as the differences in callback rates of all other groups relative to 'female: mother teacher, father unskilled worker'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Table 13

Treatment effects of gender and parental occupation by types.

	est	boot se	raw pval	adj pval
Female-dominated apprenticeship types				
female: mother teacher, father unskilled worker (mean)	0.120	0.067	0.073	
female: mother teacher, father intermediate	0.137	0.075	0.066	0.330
female: mother teacher, father professor	0.315	0.127	0.013	0.047
female: mother home, father unskilled worker	0.065	0.101	0.517	0.832
female: mother home, father intermediate	0.093	0.075	0.218	0.695
female: mother home, father professor	0.213	0.108	0.049	0.263
male: mother teacher, father unskilled worker	0.192	0.137	0.159	0.298
male: mother teacher, father intermediate	0.027	0.075	0.718	0.928
male: mother teacher, father professor	−0.072	0.080	0.368	0.832
male: mother home, father unskilled worker	0.184	0.117	0.117	0.298
male: mother home, father intermediate	0.070	0.071	0.323	0.832
male: mother home, father professor	0.005	0.096	0.958	0.955
number of observations	718			
Gender neutral apprenticeship types				
female: mother teacher, father unskilled worker (mean)	0.182	0.083	0.028	
female: mother teacher, father intermediate	0.179	0.089	0.045	0.381
female: mother teacher, father professor	0.232	0.125	0.063	0.209
female: mother home, father unskilled worker	0.161	0.116	0.163	0.490
female: mother home, father intermediate	0.093	0.085	0.271	0.816
female: mother home, father professor	0.218	0.121	0.070	0.236
male: mother teacher, father unskilled worker	0.040	0.114	0.724	0.837
male: mother teacher, father intermediate	0.131	0.089	0.138	0.617
male: mother teacher, father professor	0.061	0.108	0.576	0.837
male: mother home, father unskilled worker	0.077	0.116	0.503	0.837
male: mother home, father intermediate	0.086	0.088	0.328	0.816
male: mother home, father professor	0.096	0.112	0.394	0.816
number of observations	964			
Male-dominated apprenticeship types				
female: mother teacher, father unskilled worker (mean)	0.238	0.068	0.000	
female: mother teacher, father intermediate	0.098	0.073	0.182	0.760
female: mother teacher, father professor	0.156	0.108	0.148	0.486
female: mother home, father unskilled worker	0.070	0.101	0.490	0.849
female: mother home, father intermediate	0.088	0.069	0.206	0.825
female: mother home, father professor	0.194	0.105	0.065	0.277
male: mother teacher, father unskilled worker	0.020	0.105	0.849	0.864
male: mother teacher, father intermediate	0.084	0.073	0.250	0.825
male: mother teacher, father professor	0.140	0.104	0.179	0.538
male: mother home, father unskilled worker	0.045	0.096	0.643	0.864
male: mother home, father intermediate	0.064	0.073	0.377	0.849
male: mother home, father professor	0.088	0.097	0.364	0.825
number of observations	1246			

Note: estimates of (differences in) callback rates for tier 1 and 2 vs. tier 3, without control variables. 'est' provides the callback rate for the group 'female: mother teacher, father unskilled worker', as well as the differences in callback rates of all other groups relative to 'female: mother teacher, father unskilled worker'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Table 14

Treatment effects of gender and parental occupation by size.

	est	boot se	raw pval	adj pval
Up to 50 employees (estimated)				
female: mother teacher, father unskilled worker (mean)	0.167	0.047	0.000	
female: mother teacher, father intermediate	0.144	0.050	0.004	0.147
female: mother teacher, father professor	0.267	0.081	0.001	0.001
female: mother home, father unskilled worker	0.048	0.067	0.477	0.756

Table 14 (*continued*).

	est	boot se	raw pval	adj pval
female: mother home, father intermediate	0.085	0.049	0.083	0.583
female: mother home, father professor	0.167	0.067	0.013	0.070
male: mother teacher, father unskilled worker	0.059	0.070	0.398	0.756
male: mother teacher, father intermediate	0.092	0.052	0.077	0.526
male: mother teacher, father professor	0.009	0.070	0.901	0.899
male: mother home, father unskilled worker	0.119	0.070	0.091	0.306
male: mother home, father intermediate	0.067	0.049	0.175	0.700
male: mother home, father professor	0.080	0.068	0.239	0.625
number of observations	2092			

More than 50 employees (estimated)

	est	boot se	raw pval	adj pval
female: mother teacher, father unskilled worker (mean)	0.261	0.092	0.004	
female: mother teacher, father intermediate	0.093	0.099	0.345	0.797
female: mother teacher, father professor	0.114	0.125	0.362	0.762
female: mother home, father unskilled worker	0.191	0.125	0.126	0.525
female: mother home, father intermediate	0.085	0.094	0.371	0.819
female: mother home, father professor	0.275	0.133	0.039	0.209
male: mother teacher, father unskilled worker	0.156	0.167	0.351	0.689
male: mother teacher, father intermediate	0.084	0.102	0.411	0.819
male: mother teacher, father professor	0.121	0.125	0.333	0.762
male: mother home, father unskilled worker	0.008	0.127	0.947	0.946
male: mother home, father intermediate	0.089	0.099	0.370	0.811
male: mother home, father professor	0.042	0.124	0.733	0.925
number of observations	836			

Note: estimates of (differences in) callback rates for employers with up to 50 employees vs. more than 50 employees, without control variables. 'est' provides the callback rate for the group 'female: mother teacher, father unskilled worker', as well as the differences in callback rates of all other groups relative to 'female: mother teacher, father unskilled worker'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Table 15

Treatment effects of gender and parental occupation in September 2018.

	est	boot se	raw pval	adj pval
female: mother teacher, father unskilled worker (mean)	0.184	0.063	0.003	
female: mother teacher, father intermediate	0.118	0.067	0.080	0.515
female: mother teacher, father professor	0.248	0.103	0.015	0.058
female: mother home, father unskilled worker	0.030	0.088	0.732	0.896
female: mother home, father intermediate	0.076	0.067	0.257	0.802
female: mother home, father professor	0.159	0.099	0.108	0.350
male: mother teacher, father unskilled worker	−0.036	0.097	0.710	0.896
male: mother teacher, father intermediate	0.132	0.070	0.058	0.416
male: mother teacher, father professor	0.087	0.091	0.343	0.776
male: mother home, father unskilled worker	0.142	0.094	0.130	0.402
male: mother home, father intermediate	0.058	0.068	0.399	0.848
male: mother home, father professor	0.066	0.092	0.476	0.848
number of observations	1248			

Note: estimates of (differences in) callback rates for September 2018, without control variables. 'est' provides the callback rate for the group 'female: mother teacher, father unskilled worker', as well as the differences in callback rates of all other groups relative to 'female: mother teacher, father unskilled worker'. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing.

Table 16

Regression with control variables.

	estimate	clustered se	*t*-value	p-value
female: mother teacher, father unskilled worker (intercept)	0.323	0.089	3.617	0.000
female: mother teacher, father intermediate	0.131	0.044	2.962	0.003
female: mother teacher, father professor	0.211	0.065	3.232	0.001
female: mother home, father unskilled worker	0.090	0.058	1.544	0.123
female: mother home, father intermediate	0.084	0.042	1.996	0.046
female: mother home, father professor	0.189	0.061	3.092	0.002
male: mother teacher, father unskilled worker	0.082	0.063	1.293	0.196
male: mother teacher, father intermediate	0.079	0.044	1.769	0.077
male: mother teacher, father professor	0.052	0.059	0.887	0.375
male: mother home, father unskilled worker	0.092	0.061	1.511	0.131
male: mother home, father intermediate	0.069	0.043	1.588	0.112
male: mother home, father professor	0.050	0.058	0.866	0.387
employees: 1 to 20	−0.051	0.053	−0.951	0.342

Table 16 (*continued*).

	estimate	clustered se	*t*-value	p-value
employees: 21 to 50	−0.013	0.054	−0.239	0.811
employees: 51 to 100	0.005	0.056	0.096	0.924
employees: 101 to 250	0.034	0.060	0.567	0.571
employees: 251 to 500	−0.053	0.083	−0.639	0.523
sector: trade and wholesale	0.040	0.031	1.314	0.189
sector: manufacturing and goods	0.066	0.034	1.953	0.051
distance to city center	−0.003	0.001	−4.345	0.000
tier 1 job	−0.091	0.028	−3.204	0.001
tier 2 job	−0.184	0.031	−5.964	0.000
type: gender-neutral	−0.026	0.029	−0.900	0.368
type: female-dominated	−0.020	0.033	−0.625	0.532
city: Bern	0.201	0.034	5.868	0.000
city: Zurich	0.057	0.040	1.447	0.148
city: Basel	−0.003	0.036	−0.087	0.931
activity: international	0.012	0.046	0.256	0.798
antidiscrimination policy	−0.011	0.029	−0.388	0.698
contact: female	0.050	0.031	1.629	0.103
contact: male	0.061	0.030	2.020	0.043
day job was published or found	0.000	0.001	0.518	0.605
day of application	−0.001	0.001	−1.403	0.161

Note: Linear regression with cluster-robust standard errors, not accounting for multiple testing.

Table 17

Gender differences in callbacks by professor status and language regions.

	est	boot se	raw pval	adj pval	est	boot se	raw pval	adj pval
Father is not professor	German speaking region				French speaking region			
male (mean/intercept)	0.316				0.210			
female (diff)	0.016	0.024	0.500	0.495	0.020	0.026	0.457	0.832
number of observations	1572				983			
Father is professor	German speaking region				French speaking region			
male with prof	0.282				0.212			
female with prof	0.175	0.065	0.007	0.008	0.050	0.080	0.535	0.665
number of observations	243				130			

Note: estimates of (differences in) callback rates across gender when father is reported or not reported to be a professor by language region, without control variables. 'est' provides the callback rate or intercept for males and the difference in call back rates between females and males. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing of differences by gender, professor, and the gender–professor-interaction.

Table 18

Gender differences in callbacks by professor status and tiers.

	est	boot se	raw pval	adj pval	est	boot se	raw pval	adj pval
Father is not professor	Tiers 1 and 2				Tier 3			
male (mean/intercept)	0.234				0.371			
female (diff)	0.019	0.020	0.358	0.762	0.026	0.037	0.470	0.754
number of observations	1823				732			
Father is professor	Tiers 1 and 2				Tier 3			
male (mean/intercept)	0.236				0.321			
female (diff)	0.153	0.062	0.014	0.013	0.044	0.105	0.676	0.754
number of observations	274				99			

Note: estimates of (differences in) callback rates across gender when father is reported or not reported to be a professor by tier, without control variables. 'est' provides the callback rate or intercept for males and the difference in call back rates between females and males. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing of differences by gender, professor, and the gender–professor-interaction.

Table 19
Gender differences in callbacks by professor status and size.

	est	boot se	raw pval	adj pval	est	boot se	raw pval	adj pval
Father is not professor	Up to 50 employees				More than 50 employees			
male (mean/intercept)	0.247				0.343			
female (diff)	0.023	0.020	0.259	0.486	0.009	0.035	0.792	0.964
number of observations	1846				709			
Father is professor	Up to 50 employees				More than 50 employees			
male (mean/intercept)	0.215				0.343			
female (diff)	0.141	0.063	0.024	0.021	0.098	0.095	0.306	0.327
number of observations	246				127			

Note: estimates of (differences in) callback rates across gender when father is reported or not reported to be a professor by number of employees, without control variables. 'est' provides the callback rate or intercept for males and the difference in call back rates between females and males. 'boot se' reports bootstrap standard errors clustered at the employer level. 'raw p-val' gives the p-values not accounting for multiple hypothesis testing. 'adj p-val' provides adjusted p-values accounting for multiple hypothesis testing of differences by gender, professor, and the gender–professor-interaction.

Table 20
Neumark (2012) Method: Heteroskedastic Probit estimates for callback wrt gender.

Variables	Father is not professor	Father is professor
Estimates from basic Probit (marginal effects)	0.027 (0.018)	0.152*** (0.050)
Heteroskedastic Probit Model (marginal effects) Females (unbiased estimates)	0.030 (0.018)	0.148*** (0.059)
Marginal effect of female through level	−0.003 (0.035)	0.166 (0.109)
Marginal effect of female through variance	0.033 (0.033)	−0.019 (0.130)
Standard deviation of unobservables: Female/Male	1.178	0.901
Wald test statistic: null hypothesis that ratio of standard deviations = 1 (p-value)	0.356	0.878
Wald test statistic: null hypothesis that ratio of coefficients (of Female/Male) = 1 (p-value)	0.994	0.978
Number of observations	2555	373

p*** 0.01, p** 0.05, p* 0.10.
Notes: The variable callback measures whether an applicant was invited for an interview, to visit an assessment center or to a trial period. Standard errors are clustered at the company level. Controls are: dummy variables for company size (number of employees), dummy variables for sector of operation, distance from company to city center, dummy variables for apprenticeship tier, type of gender dominated sector, city dummies, dummy for whether the company has an international range of operations, dummy for whether an antidiscriminatory policy is explicitly stated in the company's website, and dummy variables for gender of contact person at company.

Appendix E. Constructing applications

As mentioned in the text, we consulted specialized online portals dedicated to the apprenticeship market in order to learn how real apprenticeships look like. Normally, these websites provided schematic information about which entries/details to include in both the cover letter as well as the CV. In Fig. 1, we show one such example from our files and which is still online today.[12] We very much carefully kept to this overall sensible structure. The example states first the obvious fields (name, address, phone number) but also place of birth and nationality. Of particular interest for our research was of course the information on the parents (*Eltern*), with the names and professions of both parents included. Other fields of interest included previous experience. In Switzerland, it is common for young people with ages similar to our applicants to go and work for one day at a given company. These experiences could be included directly in the CV or mentioned in the cover letter (or both). The first sentence in the example letter mentions such an activity: *Ich hatte bereits Gelegenheit, in Ihrer Unternehmung zu schnuppern.* The "applicant" is saying that she had had already the opportunity to "work" in this one-day mode at the company she is applying to. In addition, a short list of the necessary documents to complete the application (e.g. grade certificates) was also added to the letter.

In Fig. 2, we show an actual application used (with the picture deleted), with the CV on the left-hand side and the template for the cover letter on the right. When sending to an actual company, we would then add the company's specifics (name and address of the person in charge of the recruitment process, appropriate salutation, and so on) to the top of the cover letter.

---

[12] https://www.yumpu.com/de/document/view/44876248/musterbewerbung-lehrstelle-perlen-papier-ag.

Maxime Muster
Musterstrasse 111
1111 Musterstadt
041 123 45 67
079 123 45 67

Musterstadt, 14. Oktober 2008

Bewerbung für eine Lehrstelle als Kauffrau, Profil E

Sehr geehrter Herr

Ich hatte bereits Gelegenheit, in Ihrer Unternehmung zu schnuppern. Ich habe von diesem Beruf einen sehr guten Eindruck erhalten. Eigentlich alle Arbeiten, die ich in dieser Zeit erledigen konnte, haben mir gefallen. Vor allem der Kontakt mit den Kunden und das Arbeiten im Team, haben mich begeistert. Hiermit bewerbe ich mich um Ihre offene Lehrstelle als Kauffrau mit Lehrbeginn 2010.

Auch die Gespräche bei der Berufsberaterin bestätigten mir, dass der Beruf Kauffrau sehr gut zu mir passt. In der Schule habe ich gute Leistungen, ich interessiere mich vor allem für Sprachen und arbeite gerne am PC. Als Hobby betreibe ich sehr intensiv Volleyball. Ich finde darin einen guten Ausgleich zur Schule. Die weiteren Angaben über mich finden Sie im Lebenslauf.

Für Ihre wohlwollende Prüfung meiner Bewerbungsunterlagen bedanke ich mich recht herzlich und freue mich über einen baldigen Bescheid.

Freundliche Grüsse

Maxime Muster

Maxime Muster
Musterstrasse 111
1111 Musterstadt
041 123 45 67
079 123 45 67

Lebenslauf

FOTO

Name, Vorname	Muster, Max	
Adresse	Musterstrasse 111	
	1111 Musterstadt	
	041 123 45 67	
	079 123 45 67	
Geburtsdatum	01. Januar 1993	
Geburtsort	Luzern	
Nationalität	Schweiz	
Eltern	Fridolin Muster	Elektroniker
	Berta Muster	Kaufmännische Angestellte
Geschwister	Seline Muster, 1995	Primarschülerin
	Martin Muster, 1997	Primarschüler
Besuchte Schulen	2000 – 2006	Primarschule in Musterstadt
	2006 – 2008	Sekundarschule A in Musterstadt
	Zurzeit	3. Sekundarklasse A in Musterstadt
Sprachkenntnisse	Deutsch	Muttersprache
	Französisch	im 4. Jahr
	Englisch	im 2. Jahr
Hobbies	Jungwacht, Fussball, Musik hören, Freunde treffen	
Referenzpersonen	Klassenlehrer	Samuel Imfeld
		Oberstufenschulhaus „Im Grund"
		Lerchenweg 9
		1111 Musterstadt
		Tel. Schule 041 123 99 88
		Tel. privat 041 123 55 44

Musterstadt im Oktober 2009

Fig. 1. Models of cover letter and CV (*Lebenslauf*) used to construct applications.

Lebenslauf

Personalien

Name:	Meier
Vorname:	Julia

Geburtsdatum:	12.08.2004
Nationalität:	Schweizerin
Heimatort:	Zürich
Adresse:	Sperrstrasse
	4057 Basel
Telefonnummer:	079 550
Email:	meierjul12@gmail.com
Geschwister:	David, 2015
Eltern:	Matthias und Judith Meier
Beruf Vater:	Sanitärinstallateur
Beruf Mutter:	Primarschullehrerin

PHOTO

Bildung:

2017 - 2018	Sekundarschule
2011 - 2017	Primarschule

Schnupperlehre:

2017	Köchin, Zürich

Sprachen:

Deutsch:	Muttersprache
Französisch:	Grundkenntnisse
Englisch:	Grundkenntnisse

Hobbys:

Schach spielen, Wanderungen und ins Kino gehen

Referenz:

Klassenlehrerin
Maria Keller

Mit grossem Interesse habe ich Ihre Stellenanzeige auf der Website berufsberatung.ch gelesen. Sehr gerne bewerbe ich mich jetzt bei Ihnen um diese Lehrstelle. Ich möchte Ihnen mit diesem Motivationsschreiben zeigen, welche Stärken ich habe und wieso ich die geeignete Kandidatin für Ihre Firma bin.

Ich habe bereits in Ihrem Gebiet geschnuppert. Ich habe ein paar Tage im Jahr 2017 als Köchin gearbeitet. Ich mag es besonders, wenn die Aufgaben konkret sind und wenn man seine Hände aktiv brauchen kann. Dies hat mir sehr gefallen und ich war begeistert, einige Aufgaben eigenständig erfüllen zu können.

Ich würde sehr gerne bei Ihnen die Lehre absolvieren, weil Ihr Unternehmen zu meinen Interessen passt und es mir den perfekten Start in meine Berufstätigkeit ermöglicht. Ich mag in meiner Freizeit wandern, ins Kino gehen und Schach spielen. Ich bin eine kommunikationsfreudige und teamfähige Person. Weitere persönliche Einzelheiten sind in meinem Lebenslauf dargestellt.
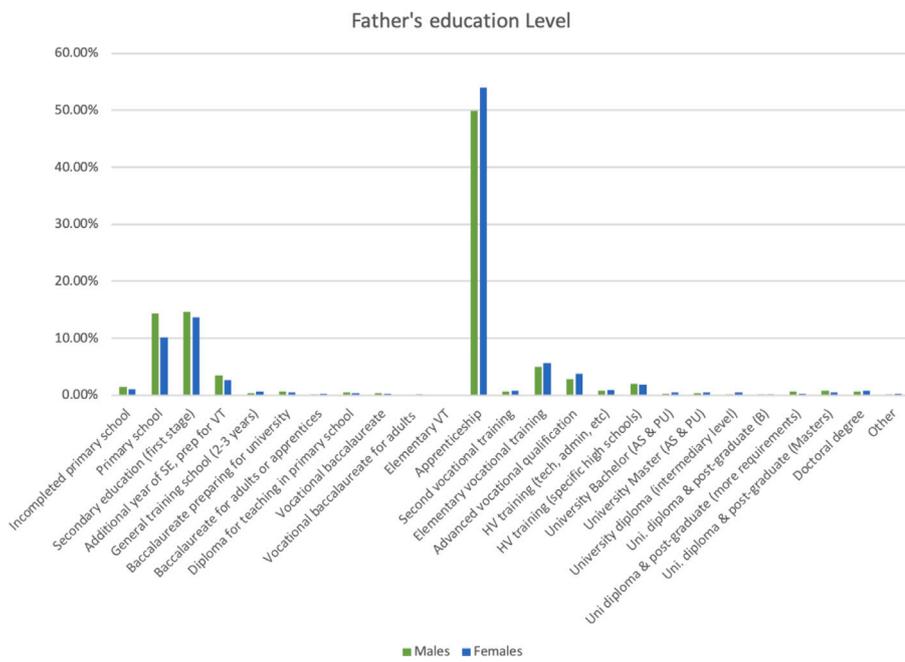
Gerne würde ich mich bei Ihnen persönlich vorstellen und freue mich auf eine Einladung zu einem Gespräch.
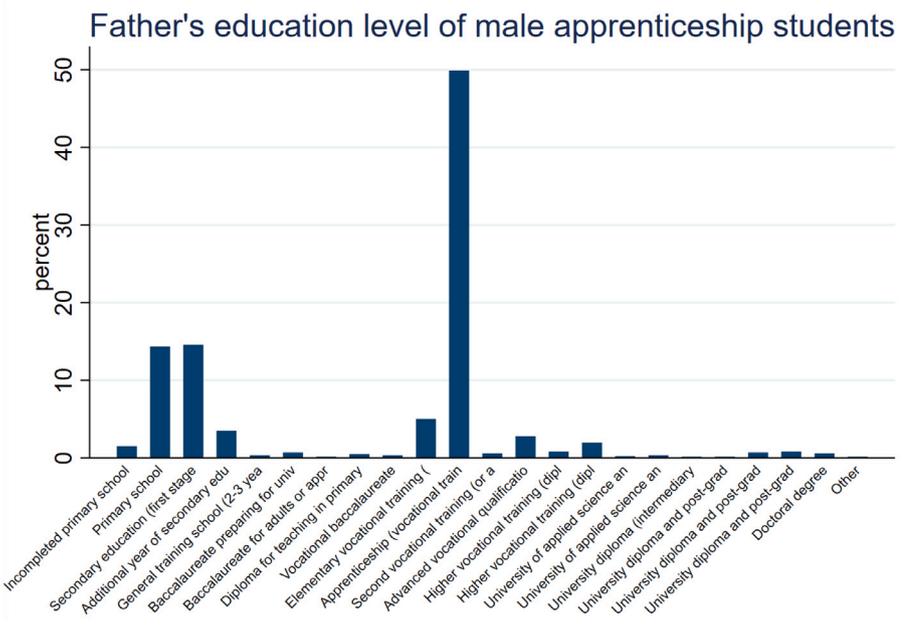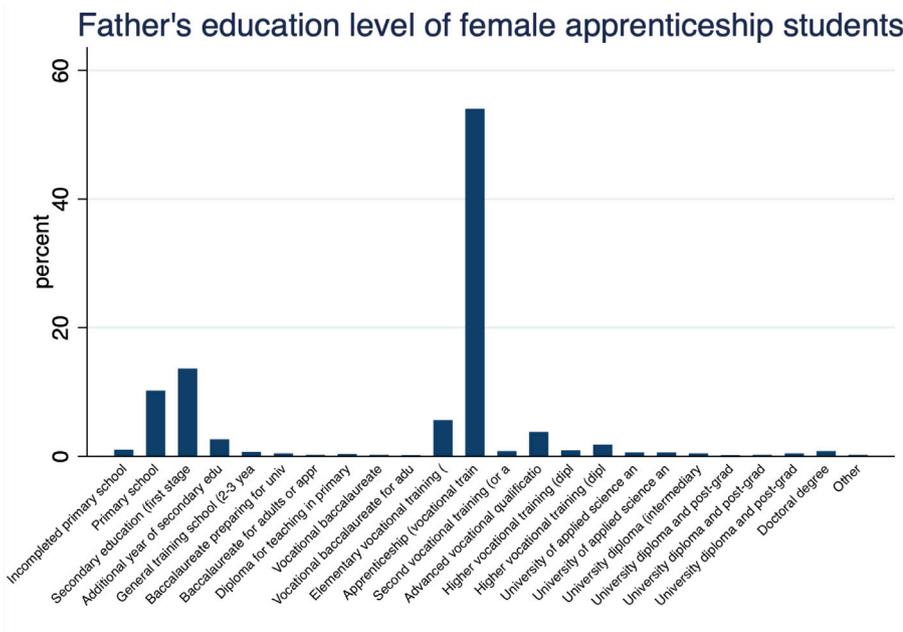
Freundliche Grüsse

Julia Meier

Fig. 2. Example of actual application constructed and used.

Appendix F. Parental professions of apprenticeship applicants

Father's education Level

Data in this section are from the from the European Social Survey (ESS) for individuals older than 25, waves 2010, 2012, 2014, 2016.

Father's education level of female apprenticeship students

Father's education level of male apprenticeship students

References

Arrow, K. (1973). The theory of discrimination. In *Discrimination in labor markets*. Princeton University Press.

Baert, S. (2018). Hiring discrimination: An overview of (almost) all correspondence experiments since 2005. In *Audit studies: behind the scenes with theory, method, and nuance, Vol. 14* (pp. 63–77). Cham: Springer.

Banton, M. (1997). The ethics of practice-testing. *Journal of Ethnic and Migration Studies, 23*, 413–420.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago Press.

Becker, G. S., & Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy, 87*(6), 1153–1189.

Becker, G. S., & Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics, 4*, 1–39.

Becker S. O., Fernandes, A., & Weichselbaumer, D. (2019). Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment. *Labor Economics, 59*, 139–152.

Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. In *Handbook of economic field experiments, Vol. 1* (pp. 309–393). Elsevier.

Björklund, A., Jäntti, M., & Nybom, M. (2017). The contribution of early-life versus labour market factors to intergenerational income persistence: A comparison of the UK and Sweden . *The Economic Journal, 127*, F71–94.

Blanden, J. (2013). Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. *Journal of Economic Surveys, 27*(1), 38–73.

Blau, F., & Kahn, L. (1996). International differences in male wage inequality: Institutions versus market forces. *Journal of Political Economy, 104*, 791–837.

Blommaert, L., Coenders, M., & van Tubergen, F. (2013). Discrimination of arabic-named applicants in the netherlands: An internet-based field experiment examining different phases in online recruitment procedures. *Social Forces, 92*, 957–982.

Bourabain, D., Verhaeghe, P.-P., & Stevens, P. A. (2020). School of choice or schools' choice? Intersectional correspondence testing on ethnic and class discrimination in the enrolment procedure to flemish kindergarten. *Race Ethnicity and Education*, 1–21.

Bursell, M. (2007). *What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process*: SULCIS reports and working papers.

Carlsson, M. (2011). Does hiring discrimination cause gender segregation in the Swedish labor market? *Feminist Economics, 17*(3), 71–102.

Carlsson, M., & Rooth, D.-O. (2012). Revealing taste-based discrimination in hiring: a correspondence testing experiment with geographic variation. *Applied Economics Letters*, *19*, 1861–1864.

Chadwick, L., & Solon, G. (2002). Intergenerational income mobility among daughters. *American Economic Review*, *92*(1), 335–344.

Chernozhukov, V., Hansen, C., & Spindler, M. (2015). Hdm: High-dimensional metrics. A Package for the Statistical Software R.

Cortes, P., & Pan, J. (2018). Occupation and gender. In *The Oxford handbook of women and the economy* (pp. 425–452). Oxford University Press.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550. http://dx.doi.org/10.1111/j.1542-4774.2011.01015.x, arXiv:https://academic.oup.com/jeea/article-pdf/9/3/522/10314305/jeea0522.pdf.

Fernandes, A., & Hevenstone, D. (2023). Income losses following first births and part-time work. Mimeo.

Fossati, F., Wilson, A., & Bonoli, G. (2020). What signals do employers use when hiring? Evidence from a survey experiment in the apprenticeship market. *European Sociological Review*, *36*(5), 760–779.

Goldin, C. (2015). A pollution theory of discrimination: Male and female differences in occupations and earnings. In *Human capital in history: the American record* (pp. 313–348). University of Chicago Press.

Guryan, J., & Charles, K. K. (2013). Taste-based or statistical discrimination: the economics of discrimination returns to its roots. *The Economic Journal*, *123*, F417–F432.

Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, *12*, 101–116.

Heckman, J. J., & Siegelman, P. (1993). The urban institute audit studies: Their methods and findings. In *Clear and convincing evidence: measurement of discrimination in America* (pp. 187–258). Washington, D.C.: The Urban Institute Press.

Hirvonen, L. H. (2008). Intergenerational earnings mobility among daughters and sons: Evidence from Sweden and a comparison with the United States. *American Journal of Economics and Sociology*, *67*(5), 777–826.

Huggett, M., Ventura, G., & Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, *101*, 2923–2954.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209.

Kübler, D., Schmid, J., & Stüber, R. (2018). Gender discrimination in hiring across occupations: a nationally-representative vignette study. *Labour Economics*, *55*, 215–229.

Kuhn, A., Schweri, J., & Wolter, S. C. (2019). *Local norms describing the role of the state and the private provision of training*: *IZA discussion paper 12159*.

Kuhn, P., Shen, K., & Zhang, S. (2020). Gender-targeted job ads in the recruitment process: Facts from a Chinese job board. *Journal of Development Economics*, *147*, Article 102531.

Lee, S. Y., & Seshadri, A. (2019). On the intergenerational transmission of economic status. *Journal of Political Economy*, *127*(2), 855–921.

Liebkind, K., Larja, L., Brylka, A. A., et al. (2016). Ethnic and gender discrimination in recruitment: Experimental evidence from Finland. *Journal of Social and Political Psychology*, *4*, 403–426.

Ludwig, J., Mullainathan, S., & Spiess, J. (2017). *Machine learning tests for effects on multiple outcomes*: *Working paper*, Harvard University.

Midtbøen, A. H. (2014). The invisible second generation? Statistical discrimination and immigrant stereotypes in employment processes in Norway. *Journal of Ethnic and Migration Studies*, *40*, 1657–1675.

Muehlemann, S., Pfeifer, H., Walden, G., Wenzelmann, F., & Wolter, S. C. (2010). The financing of apprenticeship training in the light of labor market regulations. *Labour Economics*, *17*(5), 799–809.

Mühlemann, S. (2016). The cost and benefits of work-based learning.

Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, *47*, 1128–1157.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, *56*, 799–866.

Pager, D. (2007). The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *The Annals of the American Academy of Political and Social Science*, *609*, 104–133.

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.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*, 659–661.

Riach, P. A., & Rich, J. (2002). Field experiments of discrimination in the market place. *The Economic Journal*, *112*, F480–F518.

Riach, P. A., & Rich, J. (2004). Deceptive field experiments of discrimination: are they ethical? *Kyklos*, *57*, 457–470.

Riach, P. A., & Rich, J. (2006). An experimental investigation of sexual discrimination in hiring in the english labor market. *Advances in Economic Analysis & Policy*, *5*(2), arXiv:arXiv:1112.2903v1.

Rich, J. (2014). *What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000*: *IZA discussion paper 8584*.

Romano, J. P., & Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, *100*, 94–108.

Romano, J. P., & Wolf, M. (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters*, *113*, 38–40.

State Secretariat for Education, Research, and Innovation (2013). *Relationship between public and private schools in Switzerland*: *Technical report*.

State Secretariat for Education, Research, and Innovation (2018). *Vocational and professional education and training in Switzerland: Facts and figures 2018*: *Technical report*.

Swiss Coordination Centre for Research in Education (2018). *Swiss education report 2018*: *Technical report*.

Weichselbaumer, D. (2004). Is it sex or personality? the impact of sex stereotypes on discrimination in applicant selection. *Eastern Economic Journal*, *30*(2), 159–186.

Wood, M., Hales, J., Purdon, S., Sejersen, T., & Hayllar, O. (2009). *A test for racial discrimination in recruitment practice in British cities*: *Department for work and pensions research report*.

Zschirnt, E. (2019). Evidence of hiring discrimination against the second generation: Results from a correspondence test in the Swiss labour market. *Journal of International Migration and Integration*, 1–23.