

## Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany

Hipp, Lena

Preprint / Preprint

Zeitschriftenartikel / journal article

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**

Wissenschaftszentrum Berlin für Sozialforschung (WZB)

### Empfohlene Zitierung / Suggested Citation:

Hipp, L. (2019). Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany. *European Sociological Review*, 1-28. <https://hdl.handle.net/10419/205802>

### Nutzungsbedingungen:

*Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.*

*Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.*

### Terms of use:

*This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.*

*By using this particular document, you accept the above-stated conditions of use.*

Hipp, Lena

**Article — Manuscript Version (Preprint)**

## Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany

European Sociological Review

**Provided in Cooperation with:**  
WZB Berlin Social Science Center

Suggested Citation: Hipp, Lena (2019) : Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany, European Sociological Review, ISSN 1468-2672, Oxford University Press, Oxford, Vol. forthcoming (accepted for publication: Oct. 2019), pp. 1–28

This Version is available at:  
<http://hdl.handle.net/10419/205802>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany

Lena Hipp

### Abstract

Although observational studies from many countries have consistently shown that motherhood negatively affects women's wages, experimental findings on its effect on the likelihood of being hired are less conclusive. Motherhood penalties in hiring have been reported in the US, the prototypical liberal market economy, but not in Sweden, the prototypical social-democratic welfare state. Based on a field experiment in Germany, this study examines the effects of parenthood on hiring processes in the prototypical conservative welfare state. My findings indicate that job recruitment processes indeed penalize women but not men for having children. In addition to providing theoretical explanations for why motherhood penalties in hiring are particularly likely to occur in the German context, this study also highlights several methodological and practical issues that should be considered when conducting correspondence studies to examine labor market discrimination.

Key words: gender, parenthood, discrimination, field experiment

Word count: 8,360 (excluding figures and tables)

Lena Hipp (hipp@wzb.eu), WZB Berlin Social Science Research Center Reichpietschufer 50, 10785 Berlin, Germany and University of Potsdam, Faculty of Economics and Social Sciences. August-Bebel-Straße 89, 14482 Potsdam, Germany

**Acknowledgements:** This research was funded by a grant from the German Ministry of Education and Research (Grant # 01UG1206 *The Social Psychology of Care Work*). I am indebted to Christina Botros, Nora Schneck, Jan Eilrich, Sophia Hess, Jakob Simonson, Marcel Lumkowsky, and Ferdinand Kriesche for their help in compiling the application materials and their thorough work in collecting the data. Moreover, I would like to thank Sandra Leumann, Stefan Munnes, Mareike Bünning, Friederike Molitor, Jan-Paul Heisig, as well as the editors and the anonymous reviewers from *European Sociological Review*.

## **Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany**

Sociological and economic research has repeatedly shown that employer discrimination has contributed to the persistent disadvantages that women, particularly mothers, face on the labor market (e.g., Azmat and Petrongolo 2014; England 2017). Mothers tend to earn less than childless women in most economically advanced societies, even after accounting for differences in working experience, human capital endowment, and occupational characteristics between the two groups (e.g., Budig, Misra and Boeckmann 2012, 2016; Cooke 2014; Gangl and Ziefle 2009). They also have a lower likelihood of getting hired and promoted than childless women when all other relevant characteristics are equal (e.g., Benard and Correll 2010; Correll, Benard and Paik 2007; Frodermann and Müller forthcoming; Firth 1982; Halpert, Wilson and Hickman 1993). Men, by contrast, do not seem to be penalized for parenthood in terms of wages and promotions (e.g., Correll, Benard and Paik 2007; Glauber 2008; Hodges and Budig 2010; Killewald 2012).

The occurrence of motherhood penalties in wages and promotions has been documented for many countries (e.g., Budig, Misra and Boeckmann (2012, 2016); Cooke (2014); Gangl and Ziefle (2009) on mothers' wage penalties, and Abendroth, Huffman and Treas (2014) on occupational status penalties). However, there is less experimental evidence of discrimination against mothers in the hiring process. As of today, only a couple of studies have investigated how employers react to mentions of children in job applications. Correll, Benard and Paik (2007), for example, found that in the US, men's callback rates were unaffected by whether or not they had children, but that mothers had lower callback rates than nonmothers. A more recent study by González et al. (2019) found that mothers in Spain were less likely to receive a callback than nonmothers, but these differences were not statistically significant. Bygren, Erlands-son and Gähler (2017) found for Sweden that neither men nor women experienced discrimination in the recruitment process for having children.

These cross-country differences may be an indication that discrimination is more likely to occur in liberal welfare state regimes and in settings with more traditional gender norms and family role expectations but less in social democratic welfare states and contexts with egalitarian attitudes towards gender and family responsibilities.

Mothers in the US and Spain may experience employer discrimination because they shoulder a majority of the burden of family work without being able to rely on public childcare (e.g., Bianchi et al. 2000; Hook 2017); they may therefore seem to be “riskier” hires than childless women. Mothers in Sweden, by contrast, may experience less or no discrimination both because parents divide family work more equally (e.g., Fuwa 2004; Kan, Sullivan and Gershuny 2011) and because they can rely on a well-developed system of publicly provided childcare (OECD Family Database 2017). This may reduce employers’ reservations about hiring mothers.

An alternative explanation for the divergent findings relates to differences in the timing of the studies and in their experimental design. The study from the US (Correll, Benard and Paik 2007), which found discrimination against mothers, was conducted more than ten years ago and employed a matched design; that is, two job applications were sent to the same employer. The more recent study for Sweden (Bygren, Erlands-son and Gähler 2017), which found no difference in callback rates between mothers and childless women, by contrast, employed an unmatched design; that is, just one application was sent out per employer. Moreover, it is possible that discrimination against parents was found in one setting but not in the other because the salience of parental status manipulation in the application materials differed across the two settings as government regulations in both countries prohibit the inclusion of private information on CVs.

Given the fact that discrimination at labor market entry has enduring career effects, it is crucially important that researchers and policymakers know more about the conditions under which discrimination in hiring occurs and how it can reliably be detected. I therefore conducted a field experiment on the effects of parenthood on men’s and women’s likelihood of being invited to a job interview in Germany.

Germany is an ideal context for such an experiment. First, it is a prototypical conservative welfare state that relies mainly on the family as a provider of care work (Blome 2016; Leitner 2017; Collins 2019) and has comparatively high motherhood wage penalties (Gangl and Ziefle 2009; Gash 2009; Kühhirt and Ludwig 2012). By examining whether motherhood penalties in hiring also occur outside the liberal welfare state context of the US, with its comparatively small motherhood wage penalty (e.g., Gangl and Ziefle 2009), my findings can provide some indication of how widespread discrimination against mothers in hiring actually is. Second, German job applications

usually contain detailed private information, such as the number and the age of any children, at the top of the CV (it is also common to mention on the CV that one has no children). Experimental studies on hiring practices in Germany should therefore have both higher internal and external validity than comparable studies in countries where parenthood or childlessness cannot be mentioned explicitly in the application materials.

By showing that mothers in Germany were 25 percent less likely to be invited to a job interview than nonmothers when applying to mid-level management and marketing jobs while there were no differences in invitation rates between fathers and nonfathers, my study contributes to the empirical evidence on discrimination against mothers in hiring. Moreover, it highlights several methodological issues that should be considered when studying labor market discrimination. First, field experiments that seek to detect labor market discrimination should use interview invitations rather than callbacks as outcome variables, as the latter are too diverse to unequivocally indicate employer interest, involve a high degree of coder discretion, and de facto require higher sample sizes to detect statistically significant differences between treatment and control group.<sup>1</sup> Second, my study shows that matched designs have smaller sample size requirements than unmatched designs for reliably detecting discrimination in labor markets without skill shortages and where applicants are not overqualified.

## **Gendered employment effects of parenthood**

Despite having made substantial progress in gaining access to higher education and the labor market over recent decades, women today still face considerable labor market disadvantages, particularly when they have children. Mothers are often considered to be less competent than nonmothers (Cuddy, Fiske and Glick 2004; Heilman and Okimoto 2008); they are less likely to be promoted (Abendroth, Huffman and Treas 2014; Benard and Correll 2010; Correll, Benard and Paik 2007; Firth 1982; Halpert, Wilson and Hickman 1993) and earn considerably less than childless women and men (Anderson, Binder and Krause 2003; Budig and England 2001; England, Bearak, Budig and Hodges 2016; Gangl and Ziefle 2009; Harkness and Waldfogel 2003; Joshi, Paci and Waldfogel 1999; Kricheli-Katz 2012). Having children does not, however, seem to be detrimental to men. Although parenthood may not be causally related to wage

increases, previous studies have shown that fathers on average earn higher wages and have steeper wage increases than nonfathers (e.g., Glauber 2008; Hodges and Budig 2010; Killewald 2012; Killewald and Gough 2013; Lundberg and Rose 2000, 2002).

There are several explanations for these findings. Women usually take time out of the labor market after having a child, which in turn leads to losses in both human capital and job experience (e.g., Becker 1985). Differences in wages and earnings between mothers and nonmothers may also be due to unobservable characteristics and self-selection into motherhood. Moreover, mothers may trade off higher wages for family-friendly working conditions (e.g., Kilbourne et al. 1994) and no longer work in jobs that match their qualification after childbirth (Abendroth et al. 2014). In addition to these supply-side factors, mothers may also experience discrimination by employers and be treated less favorably than nonmothers or men, even when all other characteristics are equal (e.g., Correll, Benard and Paik 2007; Gangl and Ziefle 2009).

Men, by contrast, may benefit from parenthood. They may increase their efforts at work after becoming fathers, work longer hours (Weeden, Cha and Bucca 2016), invest more in human and social capital, and attempt to move into higher-paying jobs (Pollmann-Schult 2011). In addition, fathers and nonfathers may differ on unobservable characteristics. Being the provider for a family may signal strength, loyalty, and dependability to employers (Acker 1990; Kanter 1977). Employers may hence treat childless men and fathers differently, just as they do with childless women and mothers, even when all else is equal.

Although motherhood wage penalties and fatherhood wage premiums have been documented widely in a large number of countries (Boeckmann, Misra and Budig 2015; Budig, Misra and Boeckmann 2016), the effects of parenthood on hiring practices seem to be less universal. For the US, Correll, Benard and Paik (2007) found that mothers had a lower likelihood of receiving a callback than nonmothers, while there was no difference in callback rates between fathers and nonfathers. In Sweden, by contrast, Bygren et al. (2017) did not find substantial differences in callback rates between parents and nonparents, irrespective of gender. An even more recent study conducted in Spain (González, Cortina and Rodríguez 2019) found lower callback rates for mothers than nonmothers, but the difference in callback rates was not statistically significant at conventional cut-off levels.

As all three studies employed an experimental design, explanations other than discrimination, such as differences in human capital endowments or work experience, can be excluded. Differences between countries in observed discrimination against mothers must hence be due either to differences in study design, including the experimental setup, the timing of the study, and the occupations used in the study, or to differences in welfare state characteristics and gender relations across countries.

## Theories of discrimination

Why would employers discriminate against mothers but not fathers? Theories of discrimination suggest several mechanisms for why employers may prefer members of some demographic groups to others. Theories of “statistical discrimination” (Arrow 1973; Phelps 1972) state that employers systematically hire from certain groups but not from others because pre-hire data on future productivity is unavailable or costly to obtain. Employers therefore base their hiring decisions on the (assumed) mean performance of a given demographic group or the variance in performance within a group. This reduces their risk of hiring a low-performing worker. According to this line of argumentation, employers actually behave rationally when discriminating against members of certain groups.

Theories of “status-based discrimination” (Correll, Benard and Paik 2007; Correll and Benard 2006) assume that discrimination occurs on less rational grounds. They suggest that employers apply biased standards to evaluate workers, for instance, that their expectations regarding workers’ future productivity are based on stereotypes rather than on accurate estimates. Laboratory studies on status-based discrimination in the US have shown that parenthood (like race and gender) is indeed a category that is subject to biased standards and that mothers experience discrimination even if they provide indisputable evidence of competence and commitment (Benard and Correll 2010).

Theories of “taste-based discrimination” (Becker 1957), by contrast, suggest that employers — or the clients they serve — may simply prefer members of certain (demographic) groups over others; they hence assume even more irrational grounds for discrimination than the aforementioned theories. The origins of preferences for or against certain demographic groups or certain behaviors are not clearly specified in



Becker’s (1957) original formulation of the theory. However, theories of “normative discrimination” (Benard and Correll 2010), which can be thought of as a subcategory of taste-based discrimination, assume that prevalent cultural norms systematically disadvantage certain groups.

## **Empirical predictions on hiring discrimination in Germany**

Despite the fact that the various theories attribute discrimination to different underlying mechanisms and processes, they all suggest that mothers face greater difficulties in finding a job than nonmothers. Theories of statistical and status-based discrimination suggest that employers may prefer to hire nonmothers over mothers because they assume that childless women will be more productive than mothers (for example, because they can work longer hours and are less distracted from work due to family obligations). Both theories, moreover, suggest that employers prefer fathers over non-fathers, as they may observe (or assume) that fatherhood is associated with increased work effort (e.g., Christiansen and Palkovitz 2001).

Theories of normative discrimination make similar predictions. They suggest that mothers are less likely to be invited to a job interview than nonmothers because working mothers, particularly mothers of young children, violate prevailing gender norms. Fathers, by contrast, should not be less likely to be invited to a job interview because fatherhood is not culturally incompatible with paid employment. If anything, employers may feel normatively obliged to ensure that men with children are able to fulfill their breadwinner roles and may therefore give preference to fathers over childless men (Benard and Correll 2010; Blair-Loy 2009; Collins 2019).

Based on the insights of different types of discrimination theories and the empirical findings on wage discrimination against mothers, I assume that in the German context parenthood should negatively affect the likelihood of being invited to a job interview for women but not for men. Germany has one of the largest motherhood wage penalties worldwide (Budig, Misra and Boeckmann 2012). For example, in their comparison of the US, the UK, and Germany, Gangl and Ziefle (2009) reported a net motherhood wage penalty of more than 10 percent per child for mothers in Germany but did not find such a penalty in the US or the UK. Despite the fact that public opinion in Germany on maternal employment and institutionalized childcare has become considerably more

progressive in recent years (Blome 2017), only 26.7 percent of employed women (vs. 94.5 percent of employed men) with children under the age of 10 work full-time (Keller and Haustein 2012). Taken together, I expect that mothers in Germany are less likely to be invited to a job interview than nonmothers (Hypothesis 1), whereas fathers are more likely to be invited to a job interview than nonfathers (Hypothesis 2).

## **Experimental design and materials**

To investigate the causal effect of parenthood on men’s and women’s hiring chances in Germany, I conducted the following field experiment. Between 2016 and 2017, my research team and I sent a total of 820 applications from two fictitious job seekers, either both male or both female, who differed in their parental status but were otherwise equivalent, to 410 employers in response to actual job advertisements for event manager positions.<sup>2</sup> In Germany, the job of event manager is around the middle of the occupational status scale, and there are approximately the same number of men and women working in this field (Wünsch 2008). Event managers are employed in various sectors (e.g., hospitality, banking, insurance, leisure, public service, etc.) and in organizations of different sizes.<sup>3</sup>

Conducting the study with event managers was beneficial for a variety of reasons. First, this choice minimized the risk that potential differences in discrimination between parents and nonparents would be exacerbated by occupational characteristics. Several studies have shown that labor market discrimination against women is highest in high-status occupations (e.g., Busch and Holst 2011; Cha 2013; Maume 1999). Moreover, the qualitative interviews conducted with event managers and HR managers in the hospitality and marketing sector prior to the field experiment revealed that event managers in some jobs also work on weekends and nights but that these times are predictable and well known in advance. Discrimination against parents should hence not be any more likely to occur in event management jobs than in any other white-collar occupation in the private sector. Second, conducting the experiment within an occupation that is gender-balanced and not concentrated in just one economic sector or organization type increases the generalizability of my findings. Third, the number of job openings for event managers per week on the two most widely used job search websites in Germany (monster.de and stepstone.de) is limited, usually ranging

between six and ten per week. Submitting applications for all suitable jobs allowed me to interpret differences in invitation rates between the two treatment groups in substantive terms. In other words, if, after applying for all available event manager jobs in a given time period, we observe that members of one group of applicants are more likely to be invited to job interviews than members of the other group, the group that is being discriminated against cannot compensate for this difference by sending out more applications.

The application materials consisted of a cover letter, a résumé, and a reference letter from the current employer. To create realistic and suitable applications, CVs were modeled on those of real event managers found on professional network platforms. Applicants' first names were chosen from the top-20 list in the respective age cohort (depending on the year in which the job application was sent out, the applicants were either born in 1987 or 1986). As I chose common last names for the fictitious job applicants, internet and social network searches yielded a very high number of hits and thereby reduced suspicion and the risk of detection. As is typically the case in Germany, the CVs included photos of the applicants. In order to make sure that our applicants all looked equally attractive, competent, trustworthy and so on, I pretested a large number of photos in the lab with student raters in order to select comparable pictures. Parental status was manipulated very prominently in the header of the résumé ("one child, 3 years" vs. "no children"). Moreover, the applicants stated that they were married and provided their date of birth on the CV. Making private, personal information such a prominent part of a job application is common practice in Germany and does not raise suspicion. At the time when they were applying for the new job, both fictitious candidates were still working for their first employer; they had been there for approximately five years and had been promoted around 2.5 years prior to the application being sent. The names of their current and previous employer (with whom they had completed an internship in college) were modeled after real (small) firms. The choice to use fictitious employers rather than real ones was guided by legal considerations and the fear of detection.

## Data collection

The CVs, along with the application and reference letters, were shown to and discussed with HR professionals from event management agencies and hotels and were tailored to occupational norms. Afterwards, the materials were pretested over several months in different settings and populations to ensure that differences in the design of the two job applications themselves (without the experimental condition) did not have an impact on the perceived qualifications of the applicants. When fielding the study, only the contact information of the corresponding employer, date of mailing, and title of the advertised job varied across applications. As the study ran for more than a year, the applicants' ages, dates of graduation, and so on were updated in the résumés and testimonials to ensure that the application materials were comparable over time.

Several student assistants collected the data; they knew about the purpose of the study but had to follow a strict protocol regarding the selection of job advertisements, the randomization, the dispatch of the applications, and the recording of employer responses. Full randomization was employed in the experiment using a random number generator (<https://www.random.org/>). That is, job openings were randomly assigned to either the male or female pair of applicants; likewise, the treatment was randomly assigned (a parent of a three-year-old child vs. a nonparent) as well as the application template (including photo and CV layout), and whether the application from the parent or the nonparent was sent first. In order to minimize suspicion, the two applications were sent out at a time lag of two to four days.<sup>4</sup>

The outcome variable in this study is the invitation to a job interview (see section 'Methodological Considerations' for additional elaboration on this choice). However, all types of employer responses, including missed cell phone calls, voice messages, and emails were collected, and both the content of the messages and the times of receipt were recorded. This information supplemented the archived job advertisements, the date of mailing the applications, and additional information collected on each employer (e.g., distance between the employer and the applicant's residence).

## Findings

Do employers in Germany treat parents and nonparents differently, and does this differ by gender? To answer this question, I first examined the frequencies and percentages

of applicant pairs receiving an invitation to a job interview by parental status and gender. Table 1 shows that fathers were invited to a job interview in around 14 percent of the cases and childless men in around 15 percent. Independent of whether men had a child or not, men in applicant pairs were treated equally in 86 percent of the cases (no applicant was invited to an interview in around 79 percent of the cases and both applicants in approximately 8 percent). Unequal treatment occurred in less than 13 percent of the cases: In around 7 percent of the cases, only childless men were invited and in around 6 percent of the cases, only fathers were invited. Having a child thus had no bearing on men's likelihood of being invited to a job interview.

For women, by contrast, the effects of parenthood on receiving an interview invitation differed considerably. The overall invitation rate for mothers was about 17 percent, while for childless women it was around 22 percent. Mothers and nonmothers were treated equally in 87 percent of the cases (no applicant was invited to an interview in around 74 percent of the cases and both applicants in approximately 13 percent of the cases). Unequal treatment occurred in 13 percent of the cases: Employers only invited childless women in 9 percent of the cases, whereas they only invited mothers in less than 4 percent of the cases.

To examine whether these differences are statistically significant, I conducted a McNemar's test, which is a nonparametric test to examine differences between paired proportions (Vuolo, Uggen and Lageson 2016, 2018; Riach and Rich 2002). Essentially, the test assesses whether the difference between those instances in which only one of the applicants was invited is statistically significant (i.e., differences between the discordant cells  $\pi_{10}$  and  $\pi_{01}$ ); the information from those instances in which either both applicants were invited or none was invited is used to determine the size of the difference (i.e., the information provided in the concordant cells  $\pi_{00}$  and  $\pi_{11}$ ).

Table 1: Interview invitations by gender and parental status

---

**Men**  
Overall invitation rate: 14.46% (N = 59)  
Invitation rate fathers: 14.22 %  
Invitation rate childless men: 14.72%

$\pi_{00}$ (no one invited) 78.92% (N = 161)	$\pi_{01}$ (only father invited) 6.37% (N = 13)
$\pi_{10}$ (only childless man invited) 6.86% (N = 14)	$\pi_{11}$ (both invited) 7.84% (N = 16)

---

McNemar's  $\chi^2(1) = 0.04$  Prob  $>\chi^2 = 0.8474$   
Exact McNemar significance probability = 1.0000

---

**Women**  
Overall invitation rate: 19.66% (N = 81)  
Invitation rate mothers: 16.99%  
Invitation rate childless women: 22.33%

$\pi_{00}$ (no one invited) 73.79% (N = 152)	$\pi_{01}$ (only mother invited) 3.88%* (N = 8)
$\pi_{10}$ (only childless man invited) 9.22%* (N = 19)	$\pi_{11}$ (both invited) 13.11% (N = 27)

---

McNemar's  $\chi^2(1) = 4.48$  Prob  $>\chi^2 = 0.0343$   
Exact McNemar significance probability = 0.0522

---

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

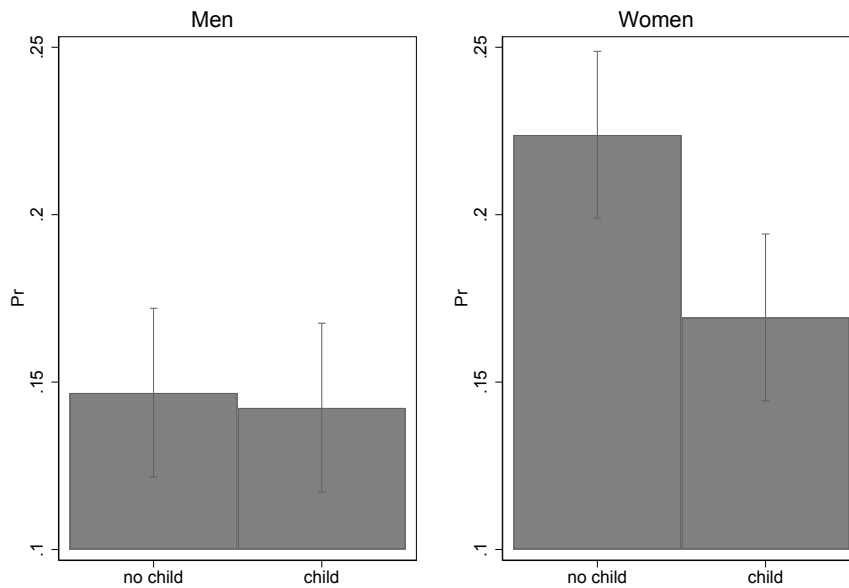
N = 820; a total of 408 applications for 204 job openings were sent by male applicants and 412 applications for 206 openings were sent by female applicants.

The results of McNemar's test show that the differences in invitations for mothers and nonmothers are statistically significant at  $p < 0.05$  (McNemar's chi-square of 4.48). As the applicants in this study were equally qualified by experimental design and as the treatment was randomly assigned, these findings suggest that parenthood does not affect how employers treat fathers compared to childless men but that it negatively affects how they treat mothers compared to childless women. Hypothesis 1, which postulated that childless women would be more likely to be invited to a job interview than mothers, is hence supported, while Hypothesis 2, which postulated that fathers would be more likely to be invited to a job interview than childless men, is empirically not supported.

As a sensitivity analysis, I also ran multivariable regressions to ensure that neither the design of the application materials nor the order in which the applications were sent out affected whether an applicant received an invitation or not (see the results

of these linear probability models with employer fixed effects and robust standard errors in Table 3 in the Appendix). The predicted probabilities based on these models are displayed in Figure 1 and confirm the findings from above. While there is no difference between fathers and nonfathers, mothers in Germany are about 25 percent (5 percentage points) less likely to be invited to a job interview than childless women. Hence, in order to be invited to an equal number of job interviews, mothers would need to submit 13 applications for every ten applications childless women submit. Given that the average number of suitable job openings for event managers in Germany at the time of data collection was just six to ten per week, this finding suggests a severe motherhood penalty in hiring in Germany when job vacancies are scarce.

Figure 1: Predicted probabilities of being invited to a job interview for men and women with and without children in the field experiment with 95% confidence intervals



N= 408

N = 412

Note: The figure displays the predicted probabilities of being invited to a job interview by gender and parental status based on the results of linear probability regressions with employer fixed effects and robust standard errors (the design of the application materials and the order in which the application materials were sent out were included as control variables).

Surprisingly, Figure 1 shows that childless men and fathers were actually much less likely to be invited than both childless women and mothers. Although gender differences in employer treatment in the hiring process were not hypothesized in this study, this finding requires some explanation and interpretation. One possible reason why women in my study were invited at higher rates than men may relate to differences in physical attractiveness between the male and female job applicants in this study. The photographs included in the résumés are a second element that varied between applications of men and women (in addition to gender). They were tested for comparability by gender, but, for obvious reasons, not between genders. In the qualitative pretests, however, test subjects considered the female applicants highly attractive, but had no particular reactions to the attractiveness of the male applicants. Thanks to their attractiveness, it might very well be that the women in my study were perceived to be a particularly good fit by the decision makers who reviewed the applications (Azmat and Petrongolo 2014). Another possible explanation refers to the fact that approximately as many women as men work as event managers in Germany (Wünsch 2008). Given that women have a higher likelihood of interrupting their careers after childbirth, more women must be hired initially to maintain this gender balance. This line of argument is backed up by findings from other audit studies that either did not find differences in employer treatment towards male and female applicants (Bertrand and Mullainathan 2004) or even found a preference for female applicants in gender-balanced occupations. For example, Riach and Rich (2006) showed that employers preferred women in accounting and programming; Carlsson and Eriksson (2019) found that women in their 30s had a higher callback rate than corresponding men; Petit (2007) found no difference between prime-aged, childless men and women applying for accounting jobs (for overviews on findings of correspondence studies on gender see Riach and Rich (2002) and Azmat and Petrongolo (2014)).

One substantive interpretation of the observed differences is that young, childless women might anticipate a slowdown of their career after childbirth and therefore make a particularly high investment in their jobs in an early stage of their career, as evidence from qualitative research suggests (Chesley 2011). Employers may have observed this behavior in the past or assume — rightly or wrongly — that young childless women will behave in this way. Another interpretation would be that (young) women are — *ceteris paribus* — less demanding than their male counterparts, in particular, when it



comes to salary negotiations, etc. (Leibbrandt and List 2014; Babcock and Laschever 2009), and that employers therefore tend to prefer women.

How does the finding of a motherhood penalty in hiring in Germany compare to findings from other countries? While my results are not directly comparable due to differences in timing, methodological choices, and other factors, it is nonetheless worth examining how large the motherhood penalty observed in Germany is compared to what has been found previously in other countries. In a study published more than ten years ago based on US data, Correll et al. (2007) conducted a matched experiment with applications for mid-level marketing positions and found that nonmothers were more than 50% more likely to receive a callback than mothers (the callback rate was 6.6 percent for childless women and 3.1 percent for mothers;  $N = 1,276$  for both male and female applicants). A more recent study by Bygren et al. (2017), which was conducted in Sweden at approximately the same time as my study, used an unmatched design to examine hiring discrimination across different occupations. In their pooled sample of 14 occupations, they found no differences between mothers and nonmothers with regard to the likelihood of callbacks in response to a job application ( $N = 2,144$  for both male and female applicants, ranging from 85 to 233 per occupational subgroup). In fact, their data showed that only assistant nurses who were mothers (invitation rate of 22 percent) were less likely to be called back by employers than nonmothers in the same occupation (60 percent); they did not detect any other statistically significant differences between mothers and nonmothers in any of the other occupations that they investigated.

Although a cross-country comparative study based on the same occupation(s) and using the same experimental design would be needed to adequately determine whether motherhood penalties in hiring were indeed contingent on country characteristics, it is plausible that mothers in Sweden are indeed less likely to experience employer discrimination than mothers in Germany or the US. Sweden is characterized by progressive gender norms and work-family policies that seek to promote a more equal distribution of childcare duties between men and women (e.g., universal childcare, moderate working hours, and incentives for men to also take family leave (e.g., Fuwa 2004; Kan, Sullivan and Gershuny 2011; OECD 2017)). This should reduce the likelihood that employers discriminate against mothers for both productivity-related and normative reasons.

## Limitations and avenues for future research

As is the case with all research, this study is limited in several ways. I therefore would like to address a few questions that may inspire future research. First, how generalizable are the findings of this study? In my experimental design, I used job applications of fictitious men and women who can be considered to be 'average' on many dimensions. Job applications were sent out to openings for the position of event managers, a gender-balanced, medium-status occupation with employees in a broad range of organizations and economic sectors (Wünsch 2008). Although these characteristics apply to a large number of jobs in Germany, my study does not allow me to draw conclusions about discrimination against mothers in general (see Heckman 1998; Neumark 2012 for a more general critique of audit studies). Replications conducted with other occupations are hence necessary. Motherhood penalties in hiring, for example, may be less pronounced in female-dominated occupations than in gender-balanced or male-dominated occupations and more likely to occur in high-status than medium status-jobs. Differences in number of job vacancies should also affect hiring discrimination. Moreover, motherhood penalties may vary with the qualifications, number of children, or children's age.

Second, how relevant is employer discrimination against mothers at the job interview stage? Although I investigated the impact of being a parent on the receipt of actual job invitations rather than callbacks, I cannot draw conclusions on discrimination against mothers at the job offer stage (e.g., Azmat and Petrongolo 2014). Researchers may consider designing experiments in which they show records of job interviews to HR managers to investigate hiring discrimination at this final stage of the application process.

Third, how can we explain discrimination against mothers in the recruitment process? Future research should examine to what degree different theoretical accounts and contextual factors can explain employer discrimination in the hiring process. Are mothers (in Germany) perceived to be less productive or committed than nonmothers (statistical and status-based discrimination)? Or do employers discriminate against mothers on normative grounds (normative and taste-based discrimination)? Answers to these questions may be found by conducting in-depth qualitative research with HR managers and other professionals involved in hiring decisions. To explore the

relevance of contextual factors, moreover, future correspondence studies should use the same experimental materials and design to assess discrimination in hiring across different policy and cultural contexts.

## Methodological considerations

Before concluding the paper, I would like to discuss a couple of methodological choices made in this study and the implications of such choices for studying labor market discrimination. By doing so, I can supplement the insights generated by the growing body of research using field experiments in economics and sociology (e.g., Bertrand and Duflo 2017; Gaddis 2018; Rich 2014 for recent overviews and practical insights on field experiments).

First, I used the invitation to a job interview as the outcome variable in this study rather than a callback, which has been used in most correspondence studies on hiring discrimination (e.g., Bertrand and Mullainathan 2004; Bygren, Erlandsson and Gähler 2017; Correll, Benard and Paik 2007; Kaas 2011; Weisshaar 2018). Although an invitation to an interview does not imply that a candidate will be offered employment, it is nonetheless an unequivocal signal of the employer’s interest in a specific candidate (see Riach and Rich 2002; González, Cortina and Rodríguez 2019 on different ways to measure employer interest in a job candidate).

Callbacks, by contrast, are not necessarily a clear sign of employer interest and require researchers to make difficult coding decisions. Are all types of callback equal in quality and seriousness? How should different types of callbacks for each of the two candidates be coded? While this issue can certainly be overcome by preregistering the experiment, it requires that all sorts of outcomes and imponderables in employer responses are anticipated before fielding the experiment (Nosek et al. 2018 for a more general discussion on the topic)<sup>5</sup>. Callbacks usually include any type of contact between the employer and the applicant, i.e., (automated) requests for further documents, questions about time availability in addition to invitations, etc.<sup>6</sup> As callback rates therefore tend to be higher than invitation rates, studies using callbacks as the outcome variable require a higher N to achieve the same statistical power as studies using invitations as the outcome variable, at least as long as the pattern in callbacks does not deviate from the pattern in invitations (see Vuolo, Uggen and Lageson 2016,

2018 for a discussion on sample size and power in different experimental designs). Table 2, which provides the information on callbacks by gender and parental status (and hence replicates the analysis from above), illustrates this point. While there are again fewer instances in which only mothers compared to only nonmothers received a callback, this difference is small and not statistically significant.

Table 2: Callbacks by gender and parental status

---

**Men**  
Overall callback rate: 21.81% (N = 89)  
Callback rate fathers: 21.08%  
Callback rate childless men: 22.55%

$\pi_{00}$ (no one called back) 70.10% (N = 143)	$\pi_{01}$ (only father called back) 7.35% (N = 15)
$\pi_{10}$ (only childless man called back) 8.82% (N = 18)	$\pi_{11}$ (both called back) 13.73% (N = 28)

---

McNemar's  $\chi^2(1) = 0.27$  Prob >  $\chi^2 = 0.6015$   
Exact McNemar significance probability = 0.7283

---

**Women**  
Overall callback rate: 28.88% (N = 119)  
Callback rate mothers: 28.16%  
Callback rate childless women: 29.61%

$\pi_{00}$ (no one called back) 64.08% (N = 132)	$\pi_{01}$ (only mother called back) 6.31%* (N = 13)
$\pi_{10}$ (only childless woman called back) 7.77% (16)	$\pi_{11}$ (both called back) 21.84% (N = 45)

---

McNemar's  $\chi^2(1) = 0.31$  Prob >  $\chi^2 = 0.5775$   
Exact McNemar significance probability = 0.7111

---

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

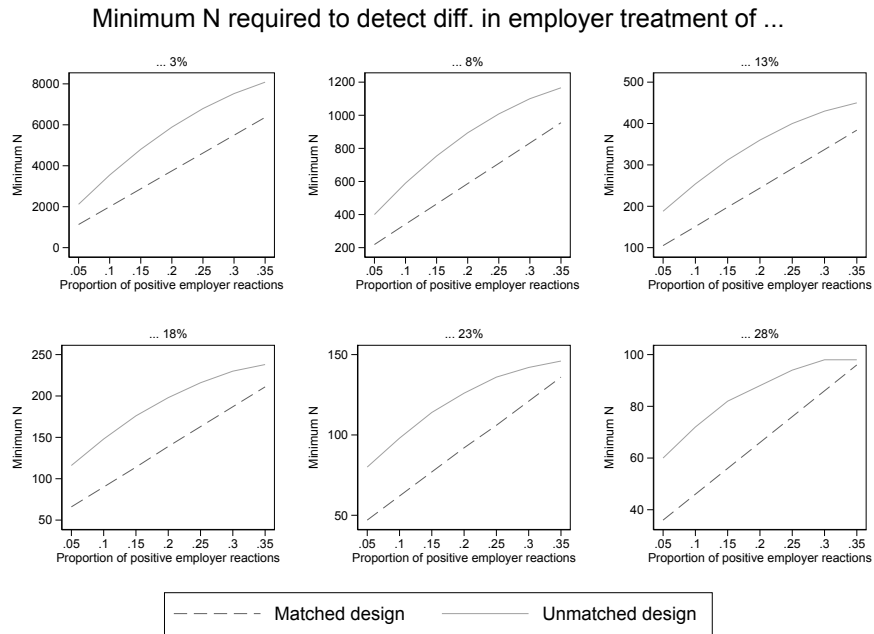
N = 820; a total of 408 applications for 204 job openings were sent by male applicants and 412 applications for 206 openings were sent by female applicants.

This null finding can be interpreted in both substantive and methodological terms. Asking for additional information and trying to learn more about applicants is less costly for employers than actually inviting applicants. It therefore comes as little surprise that employers treat mothers and nonmothers more equally in this pre-screening stage than in the invitation stage. However, this null finding may well be explained by the combination of small differences in callback rates and the small sample size, which makes it difficult to detect a difference between treatment and control group.

This consideration leads me to a second methodological issue. This study used a “matched design”, i.e., two applications were sent to the same employer. In addition to the fact that hiring decisions are usually based on comparisons between different applicants and that a matched design is therefore “more naturally aligned with most theoretical mindsets” (Charness, Gneezy and Kuhn 2012)<sup>7</sup>, a methodological consideration guided this choice. Matched designs de facto require lower sample sizes to study labor market discrimination than unmatched designs when there is no skill shortage and applicants are not overqualified. They are therefore less prone to type-II errors. This is because the required sample size to detect whether differences in employer reactions to applicants from the two groups is statistically significant at a reasonable level ( $\beta$  between 0.1 and 0.2) is contingent on both the size of this difference and the overall invitation rate (Vuolo, Uggen and Lageson 2016, 2018). Figure 2 illustrates this point by displaying sample size requirements depending on the overall callback/invitation rate and the difference in callback/invitations between the treatment and the control group.

The first panes in Figure 2 illustrate that relatively small differences in positive employer reactions between the treatment and the control group (e.g., three percentage points or less) require large sample sizes in both matched and unmatched designs. However, in such situations, sample size requirements are even higher in unmatched than in matched designs. The panes presented in the second row, moreover, show that sample size requirements between the matched and unmatched designs become smaller (and potentially negligible) with increasing rates of positive employer reactions and increasing differences between the treatment and the control group. This suggests that researchers studying discrimination should opt for matched designs to meet sample size requirements when using job applications that are neither too “bad” nor too “good” in labor markets with neither skill shortage nor high unemployment. In addition to cost considerations for the researcher and funding agency, the lower sample size requirements of matched designs in a labor market setting also have ethical implications. Sending fewer fictitious applications reduces the overall time burdens and costs imposed on employers (although of course not on the individual employer reviewing the applications)<sup>8</sup>.

Figure 2: Sample size requirement of matched and unmatched designs to detect discrimination



## Conclusion

This study examined discrimination against parents in the hiring process in Germany based on the results of a field experiment. Applications from either two fictitious male or two fictitious female job candidates were sent out to employers that had advertised job openings for event managers. The main finding of my study is that fathers were not any less likely to be invited to a job interview than nonfathers, but that mothers were 25 percent less likely to be invited than nonmothers. This means that the women who mentioned a child in their résumé had to write about one-third more applications to be invited to a job interview than the women who stated that they were childless.

This finding points to the potentially severe and long-lasting disadvantages of motherhood on the labor market. The principle of equal employment opportunities is the sine qua non of employees' wage scales and career development plans. Being offered a job interview is only the first hurdle that job applicants have to overcome to eventually be hired. When mothers have a harder time overcoming this first hurdle

than nonmothers, they may actually apply for and accept jobs that do not match their qualifications or pay adequate wages. In contrast to the predictions of compensating differentials, which suggest that mothers prefer jobs that seem to be more compatible with family obligations even if these jobs pay lower wages (e.g., Kilbourne, Farkas, Beron, Weir and England 1994), it may well be that mothers have, de facto, fewer job offers to choose from.

One implication of my study is that motherhood penalties are by no means limited to liberal, laissez-faire welfare state settings where gender inequalities tend to flourish (Cooke and Baxter 2010), as is the case for the US, where Correll et al. (2007) observed, using similar experimental design, motherhood penalties in hiring in a comparable occupation and a similar experimental design more than ten years ago. Motherhood penalties in hiring seem to also occur in conservative welfare state regimes, like Germany, that (still) rely on a male breadwinner, female caregiver model. Despite major policy efforts to advance gender equality in Germany, such as the expansion of publicly funded childcare for young children and the reform of the country's parental leave system, gender inequalities induced by parenthood still persist, as this study as well as others suggest.

My study also offers practical insights for experimental research on labor market discrimination. First, invitations rather than callbacks should be used as the outcome variable, since they provide a robust indication of employer interest in a specific applicant and reduce sample size requirements. Second, using a matched design allows researchers not only to examine the differences between job seekers (parents vs. nonparents in the case of this study), as unmatched designs do, but also to compare which applicant was preferred when only one applicant was invited. Findings from matched designs hinge less on the number or quality of other applicants for a particular job opening than findings from unmatched designs do. In most settings, moreover, matched designs have smaller sample size requirements.

In conclusion, I would like to stress one practical implication that can be derived from my study. The inclusion of parental status in job applications, which is common practice in Germany, is detrimental to gender equality. Job application systems that prohibit applicants from revealing personal information such as parental status or race would certainly be beneficial to mothers and minority applicants (at least in the first step of the recruitment process). Policy makers should therefore consider

measures that discourage the inclusion of private information that is irrelevant to job performance in job applications. In addition to information on parenthood, this includes information on race, ethnicity, and religious affiliation, as previous studies have shown. Such a small and easy-to-implement legislative change could have a major impact on labor market inequalities linked to parental status as well as many other characteristics.

## Notes

1. My study, for example, detects substantial and statistically significant differences in invitation rates between mothers and nonmothers but not in callback rates.
2. To minimize costs to the employer and avoid harm to other applicants, as soon as one of our fictitious applicants was contacted by an employer with an invitation to a job interview or a request for additional documents, etc., a research assistant sent an email saying that the applicant had just accepted another job and therefore was no longer interested in the position.
3. Event managers are responsible for planning, organizing, and marketing all sorts of events, as well as for cost calculations and public relations. There are several educational pathways into event management in Germany: either through a university degree (which was the case for the applicants in this study), vocational training, or simply through practical experience in the field.
4. A balance check revealed that the experimental condition was approximately evenly distributed between the applicants (104 vs. 100 for men and 100 vs. 96 for women) and the order in which the applications were sent out (103 vs. 101 for men and 104 vs. 102 for women). The comparable application materials and the even distribution of the experimental condition across the applicants ensured that I only measured the effect of the treatment and nothing else.
5. The decision to use invitations rather than callbacks was recorded in the grant proposal for this project and was not made during or after the data collection process.
6. In this study, the majority of the callbacks that fictitious job applicants received were requests for additional documents (particularly, diplomas or letters of recommendation), and a few were questions about the applicant's availability or willingness to move or requests to keep the application on file (see Table 4 in the Appendix).
7. Matched designs tend to be more prone to sensitization, carry-over, and learning effects than unmatched designs (Charness, Gneezy and Kuhn 2012). However, these concerns do not apply to matched designs in field experiments, in which participants are usually unaware of taking part in an experiment. This may explain why within-designs have predominantly been used in sociological correspondence analyses in the past (though not in the field of economics (see e.g. Correll, Benard and Paik 2007; Dahl and Krog 2018; Firth 1982; Kaas 2011; Peucker 2009) for sociological field studies using matched designs but Bygren, Erlandsson and Gähler (2017) as an exception).
8. However, there are also advantages of unmatched designs in the context of field experiments. These include a lower likelihood of detection, the moral argument that discrimination can only be assessed on average but not for a single employer, and the benefit that researchers do not have to bother to create comparable applications.



## References

- Abendroth, A.-K., Huffman, M. L. and Treas, J. (2014). The Parity Penalty in Life Course Perspective. *American Sociological Review*, **79**, 993–1014.
- Acker, J. (1990). Hierarchies, jobs, bodies: a theory of gendered organizations. *Gender & Society*, **4**, 139–158.
- Anderson, D. J., Binder, M. and Krause, K. (2003). The Motherhood Wage Penalty Revisited: Experience, Heterogeneity, Work Effort, and Work-Schedule Flexibility. *Industrial and Labor Relations Review*, **56**, 273–294.
- Arrow, K. J. (1973). *The Theory of Discrimination*, Princeton, N.J.: Princeton University Press, pp. 3–33.
- Azmat, G. and Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, **30**, 32–40.
- Babcock, L. and Laschever, S. (2009). *Women Don't Ask. Negotiation and the Gender Divide*, Princeton University Press.
- Becker, G. S. (1957). *The economics of discrimination*, Chicago: University of Chicago Press.
- Becker, G. S. (1985). Human Capital, Effort, and the Sexual Division of Labor. *Journal of Labor Economics*, **3**, 33–58.
- Benard, S. and Correll, S. J. (2010). Normative discrimination and the motherhood penalty. *Gender & Society*, **24**, 616–646.
- Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination, in *Handbook of Economic Field Experiments*, vol. 1, Elsevier, pp. 309–393.
- Bertrand, M. and Mullainathan, S. (2004). Are Emily and Brendan More Employable than Latoya and Tyrone? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, **94**, 991–1013.
- Bianchi, S. M., Milkie, M. A., Sayer, L. C. and Robinson, J. P. (2000). Is Anyone Doing the Housework? Trends in the Gender Division of Household Labor\*. *Social Forces*, **79**, 191–228.
- Blair-Loy, M. (2009). *Competing devotions: Career and family among women executives*, Harvard University Press.
- Blome, A. (2016). Normative Beliefs, Party Competition, and Work-Family Policy Reforms in Germany and Italy. *Comparative Politics*, **48**, 479–503.
- Blome, A. (2017). *The Politics of Work-Family Policy Reforms in Germany and Italy*, London & New York: Routledge.
- Boeckmann, I., Misra, J. and Budig, M. J. (2015). Cultural and Institutional Factors Shaping Mothers' Employment and Working Hours in Postindustrial Countries. *Social Forces*, **93**, 1301–1333.
- Budig, M. J. and England, P. (2001). The Wage Penalty for Motherhood. *American Sociological Review*, **66**, 204–225.
- Budig, M. J., Misra, J. and Boeckmann, I. (2012). The motherhood penalty in cross-national perspective: The importance of work-family policies and cultural attitudes. *Social Politics*, **19**, 163–193.

- Budig, M. J., Misra, J. and Boeckmann, I. (2016). Work–Family Policy Trade-Offs for Mothers? Unpacking the Cross-National Variation in Motherhood Earnings Penalties. *Work and Occupations*, **43**, 119–177.
- Busch, A. and Holst, E. (2011). Gender-Specific Occupational Segregation, Glass Ceiling Effects, and Earnings in Managerial Positions: Results of a Fixed Effects Model. *DIW Berlin Discussion Papers*, **1101**.
- Bygren, M., Erlandsson, A. and Gähler, M. (2017). Do Employers Prefer Fathers? Evidence from a Field Experiment Testing the Gender by Parenthood Interaction Effect on Callbacks to Job Applications. *European Sociological Review*, **33**, 337–348.
- Carlsson, M. and Eriksson, S. (2019). Age Discrimination in Hiring Decisions: Evidence from a Field Experiment in the Labor Market. *Labour Economics*.
- Cha, Y. (2013). Overwork and the Persistence of Gender Segregation in Occupations. *Gender & Society*, **27**, 158–184.
- Charness, G., Gneezy, U. and Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, **81**, 1–8.
- Chesley, N. (2011). Stay-at-home fathers and breadwinning mothers: Gender, couple dynamics, and social change. *Gender & Society*, **25**, 642–664.
- Christiansen, S. L. and Palkovitz, R. (2001). Why the “good provider” role still matters: Providing as a form of paternal involvement. *Journal of Family Issues*, **22**, 84–106.
- Collins, C. (2019). *Making Motherhood Work: How Women Manage Careers and Caregiving*, Princeton University Press.
- Cooke, L. P. (2014). Gendered Parenthood Penalties and Premiums across the Earnings Distribution in Australia, the United Kingdom, and the United States. *European Sociological Review*, **30**, 360–372.
- Cooke, L. P. and Baxter, J. (2010). “Families” in International Context: Comparing Institutional Effects Across Western Societies. *Journal of Marriage and Family*, **72**, 516–536.
- Correll, S., Benard, S. and Paik, I. (2007). Getting a Job: Is There a Motherhood Penalty? *American Journal of Sociology*, **112**, 1297–1339.
- Correll, S. J. and Benard, S. (2006). Biased Estimators? Comparing Status and Statistical Theories of Gender Discrimination. *Social Psychology of the Workplace: Advances in Group Process*, **23**, 89–116.
- Cuddy, A. J. C., Fiske, S. T. and Glick, P. (2004). When Professionals Become Mothers, Warmth Doesn’t Cut the Ice. *Journal of Social Issues*, **60**, 701–718.
- Dahl, M. and Krog, N. (2018). Experimental Evidence of Discrimination in the Labour Market: Intersections between Ethnicity, Gender, and Socio-Economic Status. *European Sociological Review*, **34**, 402–417.
- England, P. (2017). *Comparable worth: Theories and evidence*, Routledge.
- England, P., Bearak, J., Budig, M. J. and Hodges, M. J. (2016). Do Highly Paid, Highly Skilled Women Experience the Largest Motherhood Penalty? *American Sociological Review*, **81**, 1161–1189.

- Firth, M. (1982). Sex Discrimination in Job Opportunities for Women. *Sex Roles*, **8**, 891–901.
- Frodermann, C. and Müller, D. (forthcoming). Establishment Closures in Germany: The Motherhood Penalty at Job Search Durations. *European Sociological Review*.
- Fuwa, M. (2004). Macro-level Gender Inequality and the Division of Household Labor in 22 Countries. *American Sociological Review*, **69**, 751–767.
- Gaddis, S. (2018). *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, ISBN 978-3-319-71152-2.
- Gangl, M. and Ziefle, A. (2009). Motherhood, labor force behavior, and women’s careers: An empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States. *Demography*, **46**, 341–369.
- Gash, V. (2009). Sacrificing Their Careers for Their Families? An Analysis of the Penalty to Motherhood in Europe. *Social Indicators Research*, **93**, 569–586.
- Glauber, R. (2008). Race and gender in families and at work - The fatherhood wage premium. *Gender & Society*, **22**, 8–30.
- González, M. J., Cortina, C. and Rodríguez, J. (2019). The Role of Gender Stereotypes in Hiring: A Field Experiment. *European Sociological Review*, **35**, 187–204.
- Halpert, J. A., Wilson, M. L. and Hickman, J. L. (1993). Pregnancy as a Source of Bias in Performance Appraisals. *Journal of Organizational Behavior*, **14**, 649–663.
- Harkness, S. and Waldfogel, J. (2003). The family gap in pay: evidence from seven industrialised countries. *Research in Labor Economics*, **22**, 369–413.
- Heckman, J. J. (1998). Detecting Discrimination. *The Journal of Economic Perspectives*, **12**, 101–116.
- Heilman, M. E. and Okimoto, T. G. (2008). Motherhood: A potential source of bias in employment decisions. *Journal of Applied Psychology*, **93**, 189–198.
- Hodges, M. J. and Budig, M. J. (2010). Who Gets the Daddy Bonus? Organizational Hegemonic Masculinity and the Impact of Fatherhood on Earnings. *Gender & Society*, **24**, 717–745.
- Hook, J. L. (2017). Women’s Housework: New Tests of Time and Money. *Journal of Marriage and Family*, **79**, 179–198.
- Joshi, H., Paci, P. and Waldfogel, J. (1999). The wages of motherhood: better or worse? *Cambridge Journal of Economics*, **23**, 543–564.
- Kaas, C., Leo; Manger (2011). Ethnic Discrimination in Germany’s Labour Market: A Field Experiment. *German Economic Review*, **13**, 1–20.
- Kan, M. Y., Sullivan, O. and Gershuny, J. (2011). Gender Convergence in Domestic Work: Discerning the Effects of Interactional and Institutional Barriers from Large-scale Data. *Sociology*, **45**, 234–251.
- Kanter, R. M. (1977). *Men and women of the corporation*, New York: Basic Books.
- Keller, M. and Haustein, T. (2012). *Vereinbarkeit von Familie und Beruf. Ergebnisse des Mikrozensus 2011*, vol. Dezember 2012 of *Wirtschaft und Statistik*, Wiesbaden: Destatis.

- Kühhirt, M. and Ludwig, V. (2012). Domestic Work and the Wage Penalty for Motherhood in West Germany. *Journal of Marriage and Family*, **74**, 186–200.
- Kilbourne, B. S., Farkas, G., Beron, K., Weir, D. and England, P. (1994). Returns to Skill, Compensating Differentials, and Gender Bias: Effects of Occupational Characteristics on the Wages of White Women and Men. *American Journal of Sociology*, **100**, 689–719.
- Killewald, A. (2012). A Reconsideration of the Fatherhood Premium: Marriage, Coresidence, Biology, and Fathers' Wages. *American Sociological Review*, **78**, 96–116.
- Killewald, A. and Gough, M. (2013). Does Specialization Explain Marriage Penalties and Premiums? *American Sociological Review*, **78**, 477–502.
- Kricheli-Katz, T. (2012). Choice, Discrimination, and the Motherhood Penalty. *Law & Society Review*, **46**, 557–587.
- Leibbrandt, A. and List, J. A. (2014). Do women avoid salary negotiations? Evidence from a large-scale natural field experiment. *Management Science*, **61**, 2016–2024.
- Leitner, S. (2017). Reconciliation of Employment and Childcare in Austria, Germany and Iceland. Examples for Gender Equality in Family Life?, in *Gender and Family in European Economic Policy*, Springer, pp. 111–128.
- Lundberg, S. and Rose, E. (2000). Parenthood and the earnings of married men and women. *Labour Economics*, **7**, 689–710.
- Lundberg, S. and Rose, E. (2002). The Effects of Sons and Daughters on Men's Labor Supply and Wages. *The Review of Economics and Statistics*, **84**, 251–268.
- Maume, J., David J. (1999). Occupational Segregation and the Career Mobility of White Men and Women. *Social Forces*, **77**, 1433–1459.
- Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, **47**, 1128–1157.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C. and Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 2600–2606.
- OECD (2017). *OECD Employment Outlook 2017*, Paris: OECD Publishing.
- OECD Family Database (2017). OECD Family Database. PF3.4: Childcare support.
- 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.
- Peucker, M. (2009). Ethnic discrimination in the labour market - empirical evidence on a multi-dimensional phenomenon. *efms paper*, **2009**.
- Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, **62**, 659–661.
- Pollmann-Schult, M. (2011). Marriage and Earnings: Why Do Married Men Earn More than Single Men? *European Sociological Review*, **27**, 147–163.
- Riach, P. A. and Rich, J. (2002). Field experiments of discrimination in the market place. *The economic journal*, **112**, F480–F518.

- Riach, P. A. and Rich, J. (2006). An experimental investigation of sexual discrimination in hiring in the English labor market. *Advances in Economic Analysis & Policy*, **5**.
- Rich, J. (2014). What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000.
- Vuolo, M., Uggen, C. and Lageson, S. (2016). Statistical power in experimental audit studies: Cautions and calculations for matched tests with nominal outcomes. *Sociological Methods & Research*, **45**, 260–303.
- Vuolo, M., Uggen, C. and Lageson, S. (2018). *To Match or Not to Match? Statistical and Substantive Considerations in Audit Design and Analysis*, METH, 14 edn., Springer, ISBN 978-3-319-71153-9, pp. 119–141.
- Weeden, K. A., Cha, Y. and Bucca, M. (2016). Long Work Hours, Part-Time Work, and Trends in the Gender Gap in Pay, the Motherhood Wage Penalty, and the Fatherhood Wage Premium. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, **2**, 71–102.
- Weisshaar, K. (2018). From Opt Out to Blocked Out: The Challenges for Labor Market Re-entry after Family-Related Employment Lapses. *American Sociological Review*, **83**, 34–60.
- Wünsch, U. (2008). Gehaltsspiegel: Eventmanager in Deutschland. *Bad Honnef: Internationale Fachhochschule*.

**Lena Hipp** is head of the research group “Work & Care” at the WZB Berlin Social Science Center and Professor of Sociology at the University of Potsdam. Her research focuses on different aspects of social inequality, in particular work, family, gender, and organizations. Her work has been published in *Social Forces*, *Socio-Economic Review*, *Journal of Marriage and Family*, and *Work and Occupations*.

## Appendix

### Figures and Tables

Table 3: Results of linear probability regressions with employer fixed effects and robust standard errors

	Men		Women	
	<b>b</b>	<b>s.e.</b>	<b>b</b>	<b>s.e.</b>
Constant	0.160***	(0.021)	0.207***	(0.020)
Child	-0.004	(0.026)	-0.055*	(0.025)
1 <sup>st</sup> /2 <sup>nd</sup> application	-0.004	(0.026)	0.015	(0.025)
Design of application material	-0.024	(0.026)	0.019	(0.025)
$\sigma_u$	0.30		0.35	
$\sigma_e$	0.26		0.25	
N	408		412	

+ p < 0.1, \* p < 0.05, \*\* p < 0.01

Note: Standard errors are clustered by job ID to take the nonindependence of the collected data into account that results from submitting a pair of applications for the same job.

Table 4: Types of callback rates by gender, residency, and parental status

		Residency		Children	
		Munich	Hamburg	no children	children
Women	documents	5	8	6	7
	availability	3	6	4	5
	salary	0	1	1	0
	relocation	4	5	3	6
	other	11	13	11	13
	phone call	27	36	32	31
	invitation	39	42	46	35
Men	documents	8	3	6	5
	availability	2	2	2	2
	salary	0	0	0	0
	relocation	2	3	3	2
	other	13	11	12	12
	phone call	19	24	23	20
	invitation	32	27	30	21

Note: (a) The different type of callbacks, particularly missed cell phone calls and invitations, were not mutually exclusive.

(b) With any type of callback (except missed phone calls with no message), a research assistant sent an email saying that the applicant had just accepted another job and therefore was no longer interested in the position.