

Is there a rating bias of job candidates based on gender and parenthood? A laboratory experiment on hiring for an accounting job

Acta Sociologica

1–16

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00016993231204766

journals.sagepub.com/home/asj**Anni Erlandsson** 

INVEST Research Flagship Centre, University of Turku, Finland; Department of Sociology, Stockholm University, Sweden

Magnus Bygren 

Department of Sociology, Stockholm University, Sweden; Institute for Futures Studies, Sweden

Michael Gähler

Department of Sociology, Stockholm University, Sweden

Abstract

Biased practices by employers have been suggested as one possible cause for the observed gender disparities in labor market outcomes. While US-based laboratory experiments show a clear motherhood penalty in recruitment, European laboratory experiments on the topic are to our knowledge lacking. We conducted a laboratory experiment with 228 university students to study a potential gender bias in the evaluation of (fictitious) job candidates for an accounting manager position, and how recruitment decisions are made. We explore two dimensions of decision-making, that is, evaluators' individual ratings and collectively made ratings. The results show a statistically significant gender bias in job applicant ratings in favor of female applicants. Thus, female job applicants are more often than male applicants rated as the top candidates, regardless of their parental status. Also, we find no motherhood penalty in the applicant ratings. Moreover, there is a statistically significant pro-female bias in applicant ratings made by female evaluators individually and by all-female evaluation groups.

Keywords

Discrimination, employment, gender, gender bias, laboratory experiment, parenthood, recruitment

Corresponding Author:

Anni Erlandsson, INVEST Research Flagship Centre, University of Turku, Assistentinkatu 7, 20500 Turku, Finland.

Email: anni.erlandsson@sociology.su.se

Introduction

In general, men have an advantage over women in the labor market, and fathers have an advantage over mothers (Charles, 2011). Men tend to have higher wages and hold authority positions to a greater extent compared to women (Bygren and Gähler, 2012; Bygren et al., 2021; Grönlund et al., 2017). It has been suggested that discriminatory choices by employers may be one possible reason for the observed gender disparities, including motherhood penalties, in career-related outcomes (Blau and Kahn, 2017), and not least in the Swedish labor market, with its extensive rights for parental leave and reduced work hours following parenthood (e.g. Mandel and Semyonov, 2006; Gangl and Ziefle, 2009). In this study, we focus on recruitment bias based on gender, parenthood, and the combination thereof.

Previous laboratory experiments, mainly conducted in the US, find that men and women are often evaluated differently on the basis of parenthood in work-related situations, and this is to the disadvantage of mothers (Correll et al., 2007; Fuegen et al., 2004; Heilman and Okimoto, 2008). While similar European laboratory experiments are to our knowledge lacking, it is unclear whether, and to what extent, these findings are generalizable to other (laboratory) contexts. The current study takes place in Sweden and therefore provides a more family friendly Nordic context.

Studies based on other data sources suggest that we may find differences between the American and European contexts. While a US-based field experiment, that is, a correspondence test on the actual labor market, finds a motherhood penalty in recruitment (Correll et al., 2007), European field experiments do not in general find evidence of a motherhood penalty (Bygren et al., 2017; González et al., 2019; Hipp, 2020; Petit, 2007). Yet, Europe-based survey experiments show rather mixed results regarding motherhood penalty in employment (Fernandez-Lozano et al., 2020; Oesch et al., 2017). Although gender discrimination against women in recruitment has been documented internationally (e.g. Goldin and Rouse, 2000; González et al., 2019; Riach and Rich, 2006; Weichselbaumer, 2004), some recent findings even show women to be slightly advantaged over men in recruitment (Baert, 2018; Birkelund et al., 2019; Di Stasio and Larsen, 2020; Lippens et al., 2023), and the discrimination patterns vary by occupational category.

We conducted a laboratory experiment with university students in human relations or business administration, that is, potential future recruiters, in order to investigate whether men and women are rated differently when applying for a fictitious job as an accounting manager—depending on the gender of the evaluator—and examine how the (fictitious) hiring decisions are made in a laboratory setting in a European, in fact a Nordic, context. We also study whether the applicant ratings differ based on whether they are made individually or collectively and, in comparison to field experiments, we have more reliable information about the gender of the evaluator. Furthermore, a laboratory experiment allows us to observe the evaluation of job candidates in a controlled setting and may, through audit recordings of evaluation groups, give insights into the reasoning behind the recruitment decisions.

We wish to answer the following research questions: is there a gender bias in applicant ratings, that is, are the (fictitious) job candidates discriminated against based on gender? If so, does the gender bias vary with the parental status of the applicant, or with the gender of the evaluator? Moreover, we aim to explore the explicit reasoning behind the applicant ratings. Does it matter whether the recruitment decisions are made by participants individually or collectively in small groups? Also, does the gender composition of these small groups play a role in the ratings?

It appears more feasible to study gender-based in-group bias in a laboratory context rather than in the real world where problems related to data limitations and endogeneity play in (Sandberg, 2018). Yet, while there are studies relying on other methods, for example, field and survey experiments, that investigate gender bias in relation to recruiter (or evaluator) gender in employment (Booth and Leigh, 2010; Erlandsson, 2019; Fernandez-Lozano et al., 2020), European laboratory experiments studying recruitment bias based on gender and parenthood while also incorporating the gender of the evaluator are lacking, as far as we are aware.

Moreover, group interaction is important in decision-making, and the gender composition of a group is an important aspect of collective decision-making (cf. Azmat and Petrongolo, 2014, for a review of field and laboratory experiments studying gender in the labor market), which is difficult, if not impossible, to explore in field experiments. Thus, another contribution of this study is that it captures two different processes through which discrimination can take place—when hiring decisions are made individually and when hiring decisions are made collectively—and there are many real-world situations in which one or both processes are used for hiring. These processes may be expected to result in different levels of hiring discrimination because of, for example, social desirability bias, group dynamics around status, influence, and power. Yet, experimental gender research about decision-making in recruitment at the group level is scarce. Studying rating bias and the reasoning behind it is important in order to raise awareness of biased practices, and also to know more about the mechanisms behind such practices and behavior.

Theory and previous research

Homophily is a concept often used in sociology, and it refers to the tendency of individuals to sort into relationships with similar others (e.g. Kandel, 1978; McPherson et al., 2001). Also, social psychologists use the term in-group favoritism (e.g. Tajfel et al., 1971) in relation to discrimination, and individuals tend to discriminate in favor of individuals belonging to their in-group over individuals of other groups, and gender constitutes one basis for such group division (cf. Brewer and Kramer, 1985). If true, based on these ideas one would expect an own-gender bias to exist meaning that male evaluators rate male job applicants higher than female applicants while the opposite would hold for female evaluators.

This type of discrimination takes place because of a cognitive bias operating beyond the productivity of individuals of certain groups (cf. Correll and Benard, 2006), and status characteristics linked to gender and parenthood can be important in this context. A status characteristic becomes salient when it distinguishes between individual actors or is assumed to have relevance in a specific context. Status characteristics can be used (by employers or evaluators) to form expectations, for example, about individual behavior, commitment, and competence, that are in line with beliefs regarding a certain status (Wagner and Berger, 2002). For instance, status-related beliefs tend to propose men as superior to women in multiple dimensions of social life (Ridgeway, 2011), and motherhood as a salient worker characteristic is often devalued (Ridgeway and Correll, 2004a). Consequently, these types of status-related cognitive frames may lead to a motherhood penalty in job applicant ratings.

Accordingly, Benard and Correll (2010) present US-based evidence for a theory they label “normative discrimination” in relation to motherhood: normative discrimination takes place when competent and work-committed mothers are discriminated against because of employers’, possibly unconscious, belief that achievement in the labor market (especially in masculine-typed positions) indicates typical male traits, for example, being agentic. This, in turn, is inconsistent with the dominant cultural stereotype of women, as warm and caring, and high-performing mothers may thereby be expected to possess typical masculine traits and be seen as less feminine, that is, less warm and caring, making them less likable in comparison to non-mothers (Benard and Correll, 2010). This lies close to—and might provide a cognitive basis for—what in economics is termed taste discrimination, which arises from an individual’s prejudice against, or a negative “taste” for, members of a certain group (Becker, 1971). A distaste for a certain group means that a discriminating agent would be prepared to pay for avoiding this group, knowing that the group in question is as productive as any other group. While status-based discrimination resembles taste-based discrimination, their mechanisms are somewhat different, that is, a cognitive frame versus a taste, respectively. Based on these two rationales, female applicants with children would not be given lower ratings because they are assumed to be less productive than male applicants with children but because, for instance, women with children are in general considered as less suitable, and better fitted for other types of (care) work, or because of a dislike of mothers in certain kinds of (paid) work.

A theory that rather assumes group differences in mean productivity to explain discrimination is statistical discrimination (Arrow, 1973; Phelps, 1972). Statistical discrimination is based on a lack of information about the productivity of individual members of a certain group, such as men and women, or mothers and fathers. Based on statistical discrimination, one may assume recruitment discrimination against women—a pro-male bias in applicant ratings—to occur because of their lower (group level) expected productivity in comparison to men. For women of childbearing age, this may also be linked to motherhood: because mothers are assumed to bear a greater responsibility for children than fathers, being a mother can be expected to lead to lower productivity in the form of absence from work, lower commitment, and fewer working hours. Based on this, one may expect a motherhood penalty in the applicant ratings. Yet, the lower productivity of mothers, which implies costs for employers, may be attributed to all-female job candidates of childbearing age, since any of these may become a mother.

Empirical evidence

The empirical evidence on gender discrimination in the labor market is somewhat varying (Baert, 2018); discrimination patterns seem to differ depending on the country context and occupational category (cf. Birkelund et al., 2019, 2022). Whereas some field experiments on hiring detect that female subjects are negatively discriminated against (e.g. González et al., 2019 [Spain]; Petit, 2007 [France]; Riach and Rich, 2006 [the UK]; Weichselbaumer, 2004 [Austria]), others find the opposite (Berson, 2012 [France]; Booth and Leigh, 2010 [Australia]; Di Stasio and Larsen, 2020 [five European countries]; Jackson, 2009 [the UK]; Lippens et al., 2023 [a meta-study comprising many countries]), or no significant gender bias in employer callbacks (Albert et al., 2011 [Spain]; Capéau et al., 2012 [Belgium]; Kline et al., 2022 [the U.S.]). Quadlin (2018 [the US]) shows that high-achieving men receive a callback almost twice as often as high-achieving women (when achievement is indicated by the job applicant's college GPA).

While discrimination against mothers is found in the US (Correll et al., 2007), no evidence of discrimination based on parenthood, or against mothers, is found in Europe in general (Becker et al., 2019; Bygren et al., 2017; González et al., 2019; Hipp, 2020; Lippens et al., 2023; Petit, 2007) when observing which job applicants are contacted by the employer.¹ Although previous field experiments from Sweden do not show evidence of labor market discrimination based on gender (Bygren and Gähler, 2021; Carlsson, 2011), parenthood, or the combination thereof (Bygren et al., 2017), there is some evidence that male employers favor men over women, especially in gender-balanced occupations (Erlandsson, 2019).

A motherhood penalty is found in several laboratory experiments using student evaluators to explore recruitment discrimination in the US, where the institutional context differs greatly from Sweden, which is characterized by family friendly dual-earner regulations. Mothers are evaluated as less competent and have lower hiring and promotion prospects than childless women or men (Cuddy et al., 2004; Fuegen et al. 2004; Heilman and Okimoto, 2008). Correll et al. (2007) show that mothers are rated much lower on a host of measures than childless women by undergraduates, regardless of the gender of the evaluator.

While there are, to our knowledge, no previous European-based laboratory experiments investigating recruitment discrimination based on gender and parenthood, there are a few survey experiments on the topic. A Spanish survey experiment finds a motherhood premium in job promotions (Fernandez-Lozano et al., 2020) whereas a Swiss vignette study shows evidence of a motherhood penalty for female applicants for an HR assistant position (Oesch et al., 2017). A survey experiment among 239 Dutch employers finds a slight pro-female bias for a teacher position and no gender bias for a software engineer position (Mari and Luijckx, 2020); although female applicants with children are expected to have fewer working hours and lower job commitment, particularly as teachers, this does not seem to matter much for hiring decisions and offered salary.

The role of the evaluator gender

Previous experimental research presents somewhat mixed results, and inconclusive evidence, regarding the role of evaluator gender in gender discrimination. According to a meta-study of economics experiments (Lane, 2016), gender discrimination experiments show slight but statistically significant favoritism towards the opposite gender, but there is no significant difference in discriminatory behavior between females and males. Yet, according to a meta-analysis of experimental studies on employment decision-making, there is a pro-male bias in male-dominated occupations, and it is stronger by male evaluators than female evaluators (Koch et al., 2015). Foschi and Valenzuela (2012) find no bias in competence ratings based on the applicant's gender or evaluator's gender among undergraduate students.

Collective evaluations

Individual behavior can be influenced by the presence of others, that is, by group interaction, and it has been argued that the gender composition of evaluation groups can be important in collective decision-making (Azmat and Petrongolo, 2014). In contexts where individuals expect others to have gender beliefs like their own, for example, among like-minded peers, one's alternative gender beliefs, rather than the hegemonic beliefs, may become salient, and affect the behavior and evaluations of individuals (Ridgeway and Correll, 2004b). Yet, based on previous research, the influence of the gender composition of evaluation groups for collective decision-making is not quite clear, the findings are somewhat mixed and may depend on the context.

Studies find individual behavior to change when persons from the same or opposite sex are present (Antonovics et al., 2009; Gneezy and Rustichini, 2004; Ivanova-Stenzel and Kübler, 2011). Females, in comparison to males, seem to be disadvantaged in negotiation and are less keen on taking risks and engaging in competition, and some studies point to females' greater sensitivity to social signals (Azmat and Petrongolo, 2014; Bertrand, 2011). Yet, there is little experimental research on gender in relation to decision-making in recruitment at the group level, and quasi-experimental studies on the topic show somewhat mixed findings (Bagues et al., 2017; Bagues and Esteve-Volart, 2010; De Paola and Scoppa, 2015). Nevertheless, one individual rejecting dominant status norms about relative competency in a group may sway the group away from common cognitive biases (Ridgeway and Correll, 2006). This proposes that any non-stereotypical or anti-stereotypical comments made about female applicants and mothers in the group deliberations would lead to these applicants being more likely to be rated first.

Hypotheses

Because the results from previous research on employment bias are rather mixed, that is, they seem to vary by context, we do not have particularly strong expectations here. Based on recent field experiments from Europe (e.g. Becker et al., 2019; Bygren et al., 2017; Petit, 2007), we would not expect any differences in the job applicant ratings based on gender or parenthood, or the combination thereof. However, if any gender differences were to be found, based on some recent—mostly European—findings that indicate a slight female advantage in recruitment (Birkelund et al., 2019; Di Stasio and Larsen, 2020; Lippens et al., 2023), women would be expected to be advantaged over men.

Theoretically, one may expect statistical discrimination against mothers to arise because of beliefs about mothers being less productive as workers, and thereby more costly to the employer, than childless women and fathers, possibly because of expected (gender) differences in work-related characteristics and working hours following parenthood. Also, female workers (of childbearing age) in general, rather than male workers, may experience statistical discrimination on this basis, because of a potential motherhood penalty regardless of whether they have children or not. This is because they are likely either to have children or to nevertheless be perceived as “at risk” for having children. Yet, the mothers here are

presented as competent workers, and they may experience normative discrimination if they appear to deviate from the prototype of a typical mother and thereby become less likable.

Relying on the results from previous laboratory experiments, and using the normative discrimination approach as well as assumptions about mothers' productivity compared to childless women and fathers based on statistical discrimination theory, we test Hypothesis 1: there is a motherhood penalty in the applicant ratings, that is, mothers are less often rated as the top candidate than childless women, and fathers. Yet, in line with the concepts of in-group bias and homophily and the theory of taste-based discrimination, we test two more hypotheses. Hypothesis 2 states that there is a pro-male bias by male evaluators in the applicant ratings, meaning that male evaluators more often rate a male applicant as the top candidate than a female applicant. Hypothesis 3 proposes that there is a pro-female bias by female evaluators, that is, female applicants more often rate a female applicant as the top candidate than a male applicant. Finally, we apply a somewhat exploratory approach by studying whether the group evaluations (second survey) differ from the individual evaluations (first survey) in any particular respect. If anything, one might expect a stronger in-group bias if the evaluation group consists only of in-group members, that is, only female or only male evaluators, while out-group members are absent, in line with gender-based homophily. Thus, we expect to find a pro-female bias in applicant ratings among all-female evaluation groups and a pro-male bias among all-male evaluation groups.

Data and method

Participants

The data for this study come from a laboratory experiment with 228 university student participants. Each participant rated five fictitious job applications, thereby yielding a total sample of 1140 job applications. The laboratory experiment consists of two similar experiments conducted in a lecture hall at Stockholm University a few months apart, in 2016 and 2017. At the time of the study, 39 of the participants (in 2016) were attending the study program Personnel, Work, and Organization, that is, studying human relations, and 191 of the participants (in 2017) were attending an introductory undergraduate course at Stockholm Business School at Stockholm University. Thus, they are likely to work as recruiters and employers in the future.

Procedure

The participants were informed that the experiment dealt with how recruitment decisions are made, and that participation was voluntary (they were allowed to end their participation and leave the room at any time; also, they received no rewards).² They were given instructions and the experiment materials, that is, a fictitious job announcement for a position as an accounting manager in a Stockholm-based company, and five fictitious applications for the job. Management positions within economics and finance (e.g. accounting manager) are characterized by gender balance (Statistics Sweden, 2019).

First, all the participants were instructed to read the five manipulated job applications, consisting of a short resume and an application letter, and to fill out a short survey individually (the first survey). In the survey, they were asked to rank the five job applicants based on whom they would hire first, second, third, fourth, and fifth for the accounting manager position. Also, the survey included a question about the gender and the age of the participant.

Second, after completing the first survey (changes were no longer allowed), the participants were divided into groups of two or three individuals who had read the identical set of five applications. In these small groups, they were asked to discuss the qualifications of the applicants and to reach a collective decision on whom to hire for the job, that is, jointly decide on one final top candidate. The group decisions were reported in a second survey that included an open-ended question: "Who would you hire for the job as an accounting manager? Write down the name of the applicant." Also, deliberations on deciding on a top candidate within 42 groups were recorded by audit devices.

Materials

The fictitious job applications were calibrated to have equal, but not identical, qualifications, for example, relevant education and work experience for the job. Five profiles based on indicators of gender and parenthood status were rotated in the applications using a Latin Square design: each set of five applications included one applicant representing each combination of gender and parenthood, that is, one female and one male applicant who each have children and a partner, and a female and a male applicant who mention a partner but no children (the latter taken to indicate childlessness), as well as a filler profile, that is, one applicant who signals neither a partner nor children (instead, the applicant mentions friends). Thus, based on these five applicant profiles (experimental conditions), sets of five job applications were used.

Six distinctly female first names and six distinctly male first names combined with unique surnames were chosen. The age of the applicants was not specified, but their education and employment history could be taken as an indication of age. Thus, the applicants' age could be inferred to fall in the early 30s, as well as the approximate mean age at first motherhood in Stockholm (City of Stockholm, 2018).

In the job applications, we varied applicant gender, parenthood status, application number,³ name of the applicant, city of residence, and order number of the application (see Online Supplemental Material for samples of the job applications and other experiment materials). Also, the applications, surveys, and audio recordings were assigned identification numbers so that they could be linked together when analyzing the data. After the experiment, the survey responses, that is, the age and the gender of the participant and the rating of each applicant (1, 2, 3, 4, and 5) from the first survey as well as the top candidate from the second survey, were recorded.

It is a challenge to create job applications that are considered equal in terms of merits, while not being identical. Thus, professionals in management positions within audit and accounting were consulted when creating the applications. In order to assess the equivalence of the merits and the application designs, a number of colleagues and experts in financial audit and accounting were asked to read anonymized applications and to point out any differences in the qualification level of the applications. As a result, only one minor change was deemed necessary. Also, in order to further assess the job application materials and the experiment set-up, a small-scale pilot study in the form of a laboratory experiment was conducted with some junior colleagues, and subsequently, small clarifications were made to the instructions given to the participants.

Results

Table 1 shows the descriptive statistics for the data. Because each participant rated five job applications, dividing the total number of job applications (1140) by five, gives the total number of participants, which

Table 1. Descriptive statistics: number of applications by applicant gender.

	Male applicant	Female applicant
Total ($N = 1140$)	593	547
Parenthood indicated	228	228
Parenthood not indicated	365	319
Application design 1	100	128
Application design 2	101	127
Application design 3	134	94
Application design 4	118	110
Application design 5	140	88
Male evaluator	225	210
Female evaluator	368	337

is 228.⁴ The data include more female (141) than male participants (87). The age range of the participants (not shown in Table 1) is 18–46, with a mean age of 23.

Table 2 presents the proportion of applications with the highest ranking (the variable is dichotomized, i.e. ranked as the top candidate to be hired for the job versus ranked as second, third, fourth, or fifth candidate) by applicant characteristics, together with the gender ratios and the gender differences for the specified categories. We are interested in the top-ranked applicant, rather than the ranking order, because outside the laboratory employers must hire just one applicant for a position (and here it is particularly difficult to rate the applicants because of their equal qualifications). Therefore, we investigate whether a job applicant is chosen first or not. Note that the baseline chance of an application to be ranked first is 20.0%. Female applicants (23.4%) are more often than male applicants (16.9%) given the highest ranking; the gender difference of 6.5 percentage points in favor of female applicants, as well as the gender ratio of 0.72 for a male versus female applicant, are statistically significant ($p < 0.01$ for both). This contradicts expectations based on previous research, which predicts no gender difference in the applicant ratings. Another way of reporting this bias is that 127 of the female applicants got first place, but only 100 of the male applicants got first place. That is, while women made up 48% of the total applicants, 55% of the top-ranked applicants were women.

As shown in Table 2, female applicants are more often rated as the top candidate than male applicants, regardless of parenthood status. Yet, while the gender difference (8.3 percentage points) is statistically significant for men with children and women with children ($p < 0.05$), the gender difference (5.4 percentage points) between childless men and childless women does not reach statistical significance at conventional levels ($p < 0.1$). Also, there is no difference in the applicant ratings between mothers (23.7%) and childless women (23.2%) and only a very small difference (that does not reach statistical significance) in the ratings between fathers and childless men. Thus, there is no support for Hypothesis 1, that is, that there would be a motherhood penalty in the applicant ratings, as was expected based on the statistical discrimination and normative discrimination theories. In line with the expectations, there is not much of a difference in the proportion of top candidates between parents (19.5%) and childless applicants (20.3%) (not shown in Table 2).

Table 2 shows a clear and statistically significant ($p < 0.01$) pro-female bias by female evaluators, with a gender difference of 9.4 percentage points in the proportion of applicants ranked as the top candidate, whereas there is a small and not statistically significant pro-female bias by male evaluators. Therefore, Hypothesis 3, that there is a pro-female bias in applicant ratings by female evaluators, is supported, but there is no support for Hypothesis 2 concerning a pro-male bias by male evaluators.

Table A1 in the Appendix displays the conditional (fixed-effects) logistic regression estimates for the ratings of the job applicants based on the first individual survey.⁵ The dependent variable, applicant

Table 2. Proportion of job applications ranked as the top candidate by applicant gender, parental status, and evaluator gender (Survey 1).

	Number of applications	Percentage		Ratio Man versus woman	Difference Man versus woman	
		All	Male applicant (<i>N</i> = 593)			Female applicant (<i>N</i> = 547)
Total	1140	20.0	16.9	23.4	0.72**	−6.5**
Parenthood indicated	456	19.5	15.4	23.7	0.65*	−8.3*
Parenthood not indicated	684	20.3	17.8	23.2	0.77	−5.4
Male evaluator	435	20.0	19.1	21.0	0.91	−1.8
Female evaluator	705	20.0	15.5	24.9	0.62**	−9.4**

Note: For statistical tests, a two-sample test of proportions was used, and a linear regression using a log link to report risk ratios. * $p < 0.05$; ** $p < 0.01$.

rating, is dichotomous, that is, ranked as the top candidate or not. The estimated logit coefficient for *female applicants* remains positive and statistically significant ($p < 0.01$ in Models 1 and 3 and $p < 0.05$ in Models 2 and 4) throughout the models, regardless of the controls.⁶ This indicates that the finding displayed in Table 2—that female applicants are rated as the top candidate significantly more often than male applicants—is robust even when controlling for application design and applicant-related characteristics. However, the estimated logit coefficient for *female applicant × male evaluator* is negative and not statistically significant throughout the models.

Group deliberations and audio recordings

The results from the second survey, filled out collectively in small evaluation groups, are presented in Table 3 and follow a pattern similar to those of the first, individual survey. There is a clear and statistically significant ($p < 0.05$) gender difference in top applicant ratings in favor of females. Also, while a female bias appears among all-female evaluation groups (and is statistically significant at $p < 0.05$), there appears practically no gender difference in top applicant ratings among groups consisting of only male evaluators. For gender-mixed groups, the gender difference in top applicant ratings does not reach statistical significance but the estimate indicates a pro-female bias, which may mirror the fact that most gender-mixed groups are dominated by female evaluators.

Although mothers are more often rated as the top applicants than fathers, the difference is not statistically significant here, unlike in the results from the first individual survey. In the second survey, childless women (childless men) are more often ranked as the top candidates than mothers (fathers), but this difference is not statistically significant. Also, while the difference in top applicant ratings between parents and non-parents appears larger in Table 3 in comparison to Table 2, it is not statistically significant (not shown in Table 3).

A screening of the audio recordings of the group deliberations provides a few interesting examples of the reasoning behind the applicant ratings. In one case, a participant says that “Oscar has two children...it is quite good leadership training to have children” (a conversation between two male participants), but another childless male applicant is chosen as the top candidate. In other cases, parenthood status is not reflected upon, which suggests that parenthood may not be important for the applicant ratings here. In one case, a female participant expresses not only a clear female preference, but also mentions that there is a demand for female managers (a conversation between three female participants): “... We choose Malin...because of work experience, personality, and competence...and because she is a woman and women rule! ...They are needed in managerial positions.” Otherwise, the recordings

Table 3. Proportion of job applications ranked as the top candidate by applicant gender, parental status, and the gender composition of the evaluation group (Survey 2).

Percentage					Ratio	Difference
	Number of applications	All	Male applicant (N = 229)	Female applicant (N = 211)	Man versus woman	Man versus woman
Aggregated results	440	20.0	16.2	24.2	0.67*	-8.0*
Parenthood indicated	176	16.5	12.5	20.5	0.61	-8.0
Parenthood not indicated	264	22.3	18.4	26.8	0.69	-8.4
All-male evaluation group	99	20.2	20.0	20.4	0.98	-0.4
All-female evaluation group	224	20.1	15.1	25.7	0.59	-10.6*
Gender-mixed group	117	19.7	15.0	24.6	0.61	-9.6

Note: For statistical tests, a two-sample test of proportions was used, and a linear regression using a log link to report risk ratios. * $p < 0.05$.

reveal no explicit evidence of taking gender or parenthood into account. While the formal qualifications of the applicants are frequently discussed, gender and especially parenthood are rarely mentioned. Importantly, the recordings show no indication of the participants guessing or being aware of discrimination as a study interest here.

Discussion and conclusion

The current study shows no motherhood penalty in the job applicant ratings. This is in line with previous field experiments and a survey experiment from Europe, including Sweden, that show no motherhood penalty in recruitment (e.g. Bygren et al., 2017; Petit, 2007) or promotion (Fernandez-Lozano et al., 2020). Yet, unlike expectations based on most previous research, the results here show a gender bias in the ratings in favor of female applicants. This can imply that the Swedish context, with relatively high gender-egalitarian attitudes (Brandt, 2011), is to the advantage of women in comparison to a US context, even in a laboratory setting. The results suggest that discrimination on these grounds is not universal, illustrating the way national policies and culture related to family, work, and gender shape individual- and group-level decisions. Also, while Swedish law prohibits gender discrimination, it allows preferential treatment in favor of the underrepresented gender in hiring, given equal merits, in order to achieve a more gender-balanced setting (Discrimination Ombudsman, 2021).

Thus, the result that female applicants, especially when equally qualified as male applicants, are favored, or overcompensated, among university students, may be considered acceptable in Sweden. This result can be interpreted in connection to the social climate, and the participants of the laboratory experiment are likely aware of gender inequality prevailing in the labor market, that is, general gender differences disadvantaging women in Sweden. Consequently, female job applicants are favored, or overcompensated, in the current study.

At the same time, field experiments on gender discrimination show that the occupational context matters. While the specific occupation may be of importance when interpreting the findings of the current study, accounting manager can be considered as a gender-balanced position (and participants may or may not be aware of this). It is in line with some recent European field experiments to find a pro-female bias in recruitment, especially by female evaluators (Carlsson and Eriksson, 2019)—and by all-female evaluation groups—and particularly in gender-balanced occupations (Birkelund et al., 2019).

On the one hand, the findings may be interpreted against the “small wins” approach (Correll, 2017), which proposes that recruitment practices at the workplace, and even small changes such as education about gender stereotyping and bias, formalizing evaluation processes, and peer accountability among managers, can be meaningful for the evaluation of workers and for reducing gender bias in recruitment. Although the current study does not investigate changes in the workplace, peer accountability may to some extent be related to the second step of the current experiment, where the final hiring decisions are made collectively in small groups. But this, that is, making recruitment decisions collectively rather than individually, does not seem to reduce but rather increase gender bias—in favor of women. On the other hand, assuming that the “small wins” approach relies on a US context and that reducing gender bias refers to promoting female workers, one may consider the results of this study in line with the “small wins” approach: when recruitment decisions are made collectively, and evaluators may need to justify their choices to peers, then evaluators, and especially female evaluators, actively promote female applicants.

While the laboratory experiment gives a high degree of control to the researchers, it is unclear whether results produced in a laboratory setting apply to real recruitment situations in the labor market. Pager and Quillian (2005) find a difference between what employers say they will do and how they actually behave. Also, decisions made by relatively young undergraduate students may differ from decisions made by employers and recruiters in the labor market. The latter may be expected to have more work experience, as well as more experience of family formation. Thus, their beliefs about the general productivity level, work commitment, and parental leave use of male and female workers, as well as of mothers and fathers,

can be expected to differ from those of university students with limited labor market experience. However, a meta-study of economics experiments finds no difference in the tendency to discriminate between students and non-students, which suggests that the use of student participants does not produce a biased perception of the extent of discrimination by the wider population (Lane, 2016).

Yet, a social desirability bias in the applicant ratings may exist (Foschi and Valenzuela, 2012). The participants know that they are being studied, some may be primed to counter discrimination or at least be aware of discriminatory hiring behavior (particularly those enrolled in a recruitment course), and they may guess that gender or ethnic discrimination is of interest. Such knowledge could influence their responses. This may be interpreted in line with Correll's (2017) "small wins" approach in that education about gender bias, or gender inequality, matters. Thus, the participants here may consciously or unconsciously engage in gender or ethnic compensation in order not to become discriminators themselves. In line with this, Lane (2016) finds slight favoritism toward the opposite gender in gender discrimination experiments.

Although similar social desirability biases may also apply to recruitment decisions outside the laboratory setting, employers are held accountable for their decisions to a greater extent, and face consequences for hiring a job candidate. This relates to the external validity of laboratory experiments in general and is one main difference between the labor market and the laboratory setting. Hence, one may assume real employers to be less influenced by (conscious) social desirability biases than the participants of the laboratory experiment. It is important to note that labor market discrimination, and social desirability bias, can take place in different stages, such as other phases of the formal, or informal, recruitment process but also in internal promotion opportunities and wage setting. Discriminatory mechanisms may also differ between these processes.

Finally, while the results point to discrimination against men, the pro-female bias found in this study may or may not be a permanent effect. The behavior of the individuals in the study may change when they transition from the university environment to the labor market, age, and grow experienced as employers and recruiters. But if this is not the case, and their behaviors do not change, it may imply that the undergraduates studied here represent a generation more prone to strive for gender equality in the labor market, sometimes by disfavoring male applicants, than generations already in the labor market. This may be due to youth and lack of experience, or a generational context in which gender equality is widely held as an ideal.

Acknowledgements


Valuable comments from Lynn Prince Cooke, Magnus Nermo, Charlotta Stern, and three anonymous reviewers are highly appreciated.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Swedish Research Council for Health, Working Life and Welfare (Forte grant 2012-0587) and by the Strategic Research Council (SRC, established within the Academy of Finland), FLUX consortium (decision number 345130).

ORCID iDs

Anni Erlandsson  <https://orcid.org/0000-0002-9104-5049>

Magnus Bygren  <https://orcid.org/0000-0003-4398-5033>

Supplemental material

Supplemental material for this article is available online.

Notes

1. Yet, there is some evidence of a motherhood penalty when investigating the order in which the (fictitious) applicants are contacted in Spain (González et al., 2019) and when studying explicit invitations to an interview (instead of all callbacks) for an event manager position in Germany (Hipp, 2020).
2. We submitted an application to the Regional Ethical Review Board in Stockholm in 2016 (dnr. 2016/1736-31/5) prior to conducting the laboratory experiment. The board decided not to review the application: it stated that the study was not covered by the Ethical Review Act and it did not see any obstacles to carrying out the research.
3. The application number refers to the application design. Each set of applications includes five distinct application designs that differ in style, phrasing, and content (e.g. education and employment history and interests). The applications are assigned a number in order to be able to analyze the importance of application-specific features that are held constant within each application.
4. The sample included 230 participants but two of them, that is, 10 applications, were excluded because one of them did not report gender and the other reported both female and male as gender. Thus, the final sample includes 228 participants.
5. We also run the conditional logistic regression by estimating average marginal effects, and the results remain the same. In addition, we run a rank-ordered logistic regression model where the applicant ratings variable is not dichotomized, and the main results follow the same pattern as presented here.
6. To control that the results are not due to sampling variation in the application design, Models 2 and 4 include a control for application design, while Models 3 and 4 include three independent dummy variables, that is, parental status, foreign name, and city of residence, to control that the results are not due to such applicant-related characteristics.

References

- Albert R, Escot L and Fernández-Cornejo JA (2011) A field experiment to study sex and age discrimination in the Madrid labour market. *The International Journal of Human Resource Management* 22(2): 351–375.
- Antonovics K, Arcidiacono P and Walsh R (2009) The effects of gender interactions in the lab and in the field. *The Review of Economics and Statistics* 91(1): 152–162.
- Arrow K (1973) The theory of discrimination. In: Ashenfelter O and Rees A (eds) *Discrimination in Labor Markets*. Princeton, NJ: 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.
- Baert S (2018) Hiring discrimination: An overview of (almost) all correspondence experiments since 2005. In: Gaddis SM (ed.) *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. Cham, Switzerland: Springer, pp. 63–77.
- Bagues M, Sylos-Labini M and Zinovyeva N (2017) Does the gender composition of scientific committees matter? *American Economic Review* 107(4): 1207–1238.
- Bagues MF and Esteve-Volart B (2010) Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies* 77(4): 1301–1328.
- Becker GS (1971) *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.
- Becker SO, Fernandes A and Weichselbaumer D (2019) Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment. *Labour Economics* 59: 139–152.
- Benard S and Correll SJ (2010) Normative discrimination and the motherhood penalty. *Gender & Society* 24(5): 616–646.
- Berson C (2012) Does competition induce hiring equity? Documents de Travail du Centre d’Economie de la Sorbonne 12019.

- Bertrand M (2011) New perspectives on gender. In: Card D and Ashenfelter O (eds) *Handbook of Labor Economics*. Amsterdam, North-Holland: Elsevier, pp. 1543–1590.
- Birkelund GE, Janz A and Larsen EN (2019) *Do males experience hiring discrimination in female-dominated occupations? An overview of field experiments since 1996*. GEMM working paper. Available at: <https://gemm2020.eu/wp-content/uploads/2018/12/Gender-Discrimination-a-summary-1.pdf>.
- Birkelund GE, Lancee B, Larsen EN, et al. (2022) Gender discrimination in hiring: Evidence from a cross-national harmonized field experiment. *European Sociological Review* 38(3): 337–354.
- Blau FD and Kahn LM (2017) The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* 55(3): 789–865.
- Booth A and Leigh A (2010) Do employers discriminate by gender? A field experiment in female-dominated occupations. *Economics Letters* 107(2): 236–238.
- Brandt MJ (2011) Sexism and gender inequality across 57 societies. *Psychological Science* 22: 1413–1418.
- Brewer MB and Kramer RM (1985) The psychology of intergroup attitudes and behavior. *Annual Review of Psychology* 36(1): 219–243.
- 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(3): 337–348.
- Bygren M and Gähler M (2012) Family formation and men’s and women’s attainment of workplace authority. *Social Forces* 90: 795–816.
- Bygren M and Gähler M (2021) Are women discriminated against in countries with extensive family policies? A piece of the ‘welfare state paradox’ puzzle from Sweden. *Social Politics: International Studies in Gender, State & Society* 28(4): 921–947.
- Bygren M, Gähler M and Magnusson C (2021) The constant gap: parenthood premiums in Sweden 1968–2010. *Social Forces* 100(1): 137–168.
- Capéau B, Eeman L, Groenez S, et al. (2012) Two concepts of discrimination: Inequality of opportunity versus unequal treatment of equals. *Ecore Discussion Papers* 58: 127.
- Carlsson M (2011) Does hiring discrimination cause gender segregation in the Swedish labor market? *Feminist Economics* 17: 71–102.
- Carlsson M and Eriksson S (2019) In-group gender bias in hiring: Real-world evidence. *Economics Letters* 185: 108686.
- Charles M (2011) A world of difference: International trends in women’s economic status. *Annual Review of Sociology* 37: 355–371.
- City of Stockholm (2018) *Statistical Year-Book of Stockholm 2019*. Stockholm: Stockholms stad. Available at: https://start.stockholm/globalassets/start/om-stockholms-stad/utredningar-statistik-och-fakta/statistik/arsbok/arsbok_2019.pdf.
- Correll SJ (2017) SWS 2016 Feminist lecture: reducing gender biases in modern workplaces: A small wins approach to organizational change. *Gender & Society* 31(6): 725–750.
- Correll SJ and Benard S (2006) Biased estimators? Comparing status and statistical theories of gender discrimination. *Advances in Group Processes* 23: 89–116.
- Correll SJ, Benard S and Paik I (2007) Getting a job: Is there a motherhood penalty? *American Journal of Sociology* 112: 1297–1338.
- Cuddy AJC, Fiske ST and Glick P (2004) When professionals become mothers, warmth doesn’t cut the ice. *Journal of Social Issues* 60: 701–718.

- De Paola M and Scoppa V (2015) Gender discrimination and evaluators' gender: Evidence from Italian academia. *Economica* 82(325): 162–188.
- Discrimination Ombudsman (2021) Arbetslivet [Work life]. Available at: <https://www.do.se/diskriminering/diskriminering-olika-delar-samhallet/diskriminering-pa-jobbet>.
- Di Stasio V and Larsen EN (2020) The racialized and gendered workplace: Applying an intersectional lens to a field experiment on hiring discrimination in five European labor markets. *Social Psychology Quarterly* 83(3): 229–250.
- Erlandsson A (2019) Do men favor men in recruitment? A field experiment in the Swedish labor market. *Work and Occupations* 46(3): 239–264.
- Fernandez-Lozano I, González MJ, Jurado-Guerrero T, et al. (2020) The hidden cost of flexibility: A factorial survey experiment on job promotion. *European Sociological Review* 36(2): 265–283.
- Foschi M and Valenzuela J (2012) Who is the better applicant? Effects from gender, academic record, and type of decision. *Social Science Research* 41(4): 949–964.
- Fuegen K, Biernat M, Haines E, et al. (2004) Mothers and fathers in the workplace: How gender and parental status influence judgments of job-related competence. *Journal of Social Issues* 60(4): 737–754.
- 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.
- Gneezy U and Rustichini A (2004) Gender and competition at a young age. *American Economic Review* 94(2): 377–381.
- Goldin C and Rouse C (2000) Orchestrating impartiality: The impact of 'blind' auditions on female musicians. *American Economic Review* 90(4): 715–741.
- González MJ, Cortina C and Rodríguez J (2019) The role of gender stereotypes in hiring: A field experiment. *European Sociological Review* 35(2): 187–204.
- Grönlund A, Halldén K and Magnusson C (2017) A Scandinavian success story? Women's labour market outcomes in Denmark, Finland, Norway and Sweden. *Acta Sociologica* 60(2): 97–119.
- Heilman ME and Okimoto TG (2008) Motherhood: A potential source of bias in employment decisions. *Journal of Applied Psychology* 93: 189–198.
- Hipp L (2020) Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany. *European Sociological Review* 36(2): 250–264.
- Ivanova-Stenzel R and Kübler D (2011) Gender differences in team work and team competition. *Journal of Economic Psychology* 32(5): 797–808.
- Jackson M (2009) Disadvantaged through discrimination? The role of employers in social stratification. *British Journal of Sociology* 60(4): 669–692.
- Kandel DB (1978) Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology* 84(2): 427–436.
- Kline P, Rose EK and Walters CR (2022) Systemic discrimination among large US employers. *The Quarterly Journal of Economics* 137(4): 1963–2036.
- Koch AJ, D'Mello SD and Sackett PR (2015) A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology* 100(1): 128.
- Lane T (2016) Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review* 90: 375–402.
- Lippens L, Vermeiren S and Baert S (2023) The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *European Economic Review* 151: 104315.

- Mandel H and Semyonov M (2006) A welfare state paradox: State interventions and women's employment opportunities in 22 countries. *American Journal of Sociology* 111: 1910–1949.
- Mari G and Luijckx R (2020) Gender, parenthood, and hiring intentions in sex-typical jobs: Insights from a survey experiment. *Research in Social Stratification and Mobility* 65: 100464.
- McPherson M, Smith-Lovin L and Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1): 415–444.
- Oesch D, Lipps O and McDonald P (2017) The wage penalty for motherhood: Evidence on discrimination from panel data and a survey experiment for Switzerland. *Demographic Research* 37: 1793–1824.
- Pager D and Quillian L (2005) Walking the talk? What employers say versus what they do. *American Sociological Review* 70(3): 355–380.
- 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 ES (1972) The statistical theory of racism and sexism. *American Economic Review* 62(4): 659–661.
- Quadlin N (2018) The mark of a woman's record: Gender and academic performance in hiring. *American Sociological Review* 83(2): 331–360.
- Riach PA and Rich J (2006) An experimental investigation of sexual discrimination in hiring in the English labor market. *Advances in Economic Analysis and Policy* 5(2): 1–20.
- Ridgeway CL (2011) *Framed by Gender: How Gender Inequality Persists in the Modern World*. New York: Oxford University Press.
- Ridgeway CL and Correll SJ (2006) Consensus and the creation of status beliefs. *Social Forces* 85(1): 431–453.
- Ridgeway CL and Correll SJ (2004a) Motherhood as a status characteristic. *Journal of Social Issues* 60(4): 683–700.
- Ridgeway CL and Correll SJ (2004b) Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & Society* 18(4): 510–531.
- Sandberg A (2018) Competing identities: A field study of in-group bias among professional evaluators. *The Economic Journal* 128(613): 2131–2159.
- Statistics Sweden (SCB) (2019) The Swedish Occupational Register with Statistics 2017 [Yrkesregistret med yrkesstatistik 2017 Yrkesstrukturen i Sverige]. *Statistiska meddelanden AM 33 SM 1901*. Available at: https://www.scb.se/contentassets/1fe7f957920f4eaf97bddcc0270553f2/am0208_2017a01_sm_am33sm1901.pdf.
- Tajfel H, Billig MG, Bundy RP, et al. (1971) Social categorization and intergroup behaviour. *European Journal of Social Psychology* 1(2): 149–178.
- Wagner DG and Berger J (2002) Expectation states theory: An evolving research program. In: Berger J and Zelditch MJ (eds) *New Directions in Contemporary Sociological Theory*. Lanham, MD: Rowman & Littlefield, pp. 41–76.
- 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.

Author biographies

Anni Erlandsson is a Senior Researcher at the INVEST Research Flagship Centre, University of Turku, Finland, and affiliated with the Department of Sociology, the Demography Unit, Stockholm University. Her main research interests include family sociology and labor market inequality. Recent publications have appeared in *European Sociological Review*, *Work and Occupations*, and *Population Research and Policy Review*.

Magnus Bygren is a Professor of Sociology at the Department of Sociology, Stockholm University, and affiliated with the Institute for Futures Studies, Stockholm. His research centers on understanding ethnic, gender, and class inequality processes. Recent publications have appeared in *European Sociological Review*, *Social Forces*, *Social Science Research*, and *Sociology of Education*.

Michael Gähler is a Professor of Sociology at the Department of Sociology, Stockholm University. His research interests include the effects of divorce and separation for children and adults, the occurrence and extent of employer discrimination by ethnicity, gender, and parenthood, and the effects of parenthood on men's and women's careers and wages. Recent publications have appeared in *European Sociological Review*, *Social Forces*, *Social Politics*, and *Social Science Research*.

Appendix

Table A1. Conditional (fixed-effects) logistic regression of top applicant ratings (dichotomous) on the gender of the job applicant and gender of the evaluator.

	Model 1	Model 2	Model 3	Model 4
Female applicant	0.494** (0.175)	0.421* (0.178)	0.514** (0.175)	0.449* (0.179)
Male evaluator				
Female applicant × male evaluator	-0.398 (0.280)	-0.333 (0.285)	-0.413 (0.281)	-0.356 (0.286)
Controls	No	Yes	Yes	Yes
Pseudo R-squared	0.0114	0.0533	0.0218	0.0580
N job applications	1140	1140	1140	1140
Log-likelihood	-362.77758	-347.38176	-358.93787	-345.67975

* $p < 0.05$, ** $p < 0.01$.

Note: Conditional (fixed-effects) logistic regression model with standard errors in parentheses. The coefficients reported under Models 2 and 4 are conditional on the application design as a (indicator) control variable. Models 3 and 4 also include controls for parenthood, foreign name, and city of residence. Two hundred and twenty-eight participants rated five candidates each and generated a sample of 1140 job applications.