RECENT NOTES ON LABOR SCIENCE AND ORGANIZATION

Impact of Organizational Context on Gendered Recruiting Decisions among Science Engineering and Technology Professionals: An Experiment

Christina Keinert-Kisin and Sabine T. Koeszegi

Recent Notes on Labor Science and Organization (ISSN 2414-665X (online)) publishes original research in the areas of labor science, human resources, innovation, gender, organization theory and organizational behavior. As an open access journal it aims at providing current developments of the state-of-the-art. Contributions are subject to a quality control and editing procedure.

01/2015
ISSN 2414-665X (online)
edited by
Sabine T. Koeszegi & Michael Filzmoser

Impact of Organizational Context on Gendered Recruiting Decisions among Science Engineering and Technology Professionals: An Experiment

Christina Keinert-Kisin and Sabine T. Koeszegi[1]

[1] Corresponding Author: Institute of Management Science, TU Wien, Theresianumgasse 27, 1040 Vienna, email: sabine.koeszegi@tuwien.ac.at

Abstract. The purpose of this study is to systematically examine varied organizational contexts in which gender bias is expected to thrive. Discrimination against women is hypothesized to manifest itself implicitly in the assessment of suitability and potential of job seeking candidates. Data were obtained through an experiment among 296 full professors, senior scientists and students at a Science, Engineering and Technology (SET) University. Women's opportunities to be ranked for a job interview are contrasted with men's using Bradley Terry log-linear models for partial rankings and justifications for the respective ranking decisions are analyzed using content analysis. The findings show that women are ascribed significantly less relevant characteristics and skills in SET than men and are significantly less often ranked for job interviews by even experienced decision makers. Furthermore, homophilous pressures to select "socially compatible" candidates fortify discriminatory selection, while the request to respect anti-discrimination law in recruitment cannot prevent discriminating decisions. Implications of findings for organizational practice are discussed.

Keywords: gender bias, recruiting, homophily, field experiment, science engineering and technology.

1 Introduction

Women have made progress in their representation in overall managerial and professional populations, but remain underrepresented in top positions in management (Oakley, 2000; Wood, 2006), in the public sector (Kolpin & Singell, 1996; Olsson & Pringle, 2004) and in academia (Acker, 2006; Gilbert, 2007; Laabs, 1993). There is the suggestion that organizations play a key role in perpetuating inequality for their employees' career outcomes (Castilla, 2008; Fernandez & Mors, 2008; Reskin, 1993; Reskin & McBrier, 2000). While there is widespread acknowledgment that organizational practices can cause discriminatory career outcomes (England, 1992; Nelson & William, 1999; Petersen & Morgan, 1995; Petersen & Saporta, 2004), knowledge of specific organizational givens as determinants of discrimination is limited. Recent experimental research has examined discriminatory treatment of women in the field and concluded that women have a lower chance of being offered high-status or engineering positions when compared with equally qualified male competitors. As a concrete, organizational condition, meritocracy-focused

organizations were paradoxically found to place women at a disadvantage (Castilla, 2008; Fernandez & Mors, 2008; Shantz, Wright & Latham, 2011). This suggests that there is a need for more knowledge about organizational practices and characteristics that adversely affect the careers of qualified women.

In an effort to expound upon existing research, the aim of this study is to systematically examine varied organizational contexts in which gender bias is expected to thrive. Discrimination against women is hypothesized to manifest itself less explicitly than in the past, it is expected to happen on the suitability assessment level. Equally qualified male and female potentials are expected to receive unequal suitability and potential assessments that lead to unequal selection outcomes for women, but we suggest that inequality varies within the organizational context. Composition of the workforce and majority-minority relations as an organizational framework may impact suitability and potential assessments. Firstly, gender being a salient feature in application processes (King et al., 2011), perceptions of suitability of qualified female applicants are expected to be influenced by gender stereotypes for male sex-typed professions. Secondly, we propose social homogeneity in an organizational and professional context may also have an impact on perceptions of (lack of) team fit. Finally, organizations implement policies to guide individual decision-making behaviors in all areas including personnel matters. Concrete rules and guidelines against gender discrimination may mitigate the effects of homogeneity pressures and gender stereotypes.

A theory-driven, single-blind laboratory experiment was developed to test the effect of organizational context on recruiting decisions. The experiment was conducted in a respected European university in the field of Science, Engineering and Technology (SET) with approximately 3,500 employees and 20,000 students. With 296 subjects the experiment is of a large scale compared to similar laboratory experiments and also includes a substantial proportion of senior professionals (professors and senior scientists) with solid recruitment experience.

The remainder of this paper is structured as follows: In Section 2, relevant literature is reviewed and state-of-the-art research hypotheses for the empirical study are developed; Section 3 describes the laboratory experiment, the sample and the organizational settings; The detailed results of the analysis of the qualitative and quantitative experimental data are provided in Section 4, which also contains the discussion of our findings; Section 5 provides conclusions, limitations and suggestions for further research.

2 Theoretical Background and Hypotheses

There is evidence of unequal career outcomes for men and women of equal education and professional experience in firms and institutions from all (organizational) walks of life (Acker, 2006). A recent longitudinal study has matched perfect pairs of “virtual twins” of business school graduates in Austria. The pairs, who differed only by gender, were identical in academic qualification, personality type, and career ambitions at the time of their graduation. Even in the first couple of years of professional experience, women lagged behind their male “twins” in pay and career

development (Strunk et al. as cited in Mayrhofer, Meyer & Steyrer, 2007). This early career gap, controlling for qualification and personality differences, suggests organizational practices may be to blame for lack of equal opportunity for qualified and ambitious women. Because these disadvantages appear at such early career stages, family responsibilities and related part-time work cannot explain the unequal achievements. The difference in career development for highly qualified women compared to equally qualified male peers was observed in another European study for the banking sector, and it controlled for family responsibilities as well (Wood, 2006). While these studies have ruled out common explanations for an achievement gap – individual choice and the mindset for success – concrete organizational factors that may uphold and foster gender discrimination merit increased attention. In the following segment, the relevant factors for hypotheses generation are systematically analyzed.

2.1 Gender Stereotypes and Implicit Associations in a Professional Context

In the past, gender stereotypes associated with entire professions have been theoretically suggested to be a potential root cause for different evaluations for men and women regarding precisely the same trait, quality or behavior. Gender stereotypes are deeply rooted, commonly held beliefs in a given cultural context relating to characteristics, attributes and appropriate roles for men and women (Heilman, Block, Martell & Simon, 1989; Schein, 2001). Implicit stereotypes are thought to be developed from early-learned notions of masculinity and femininity. This causes a quasi-automatic association of stereotypical characteristics, abilities and roles associated with gender at a pre-conscious level, principally independent of ideology or worldview. Gender stereotypes affect management and leadership in general. Agentic traits like task and performance orientation, assertiveness, confidence and a desire to control (Sosik & Godshalk, 2000) are not only expectations of effective leadership, but also ascribed to the male, rather than the female stereotype (Eagly & Johannesen-Schmidt, 2001; Eagly, Wood & Diekman, 2000; Heilman et al., 1989; Schein, 2001; Scott & Brown, 2006). Similarly, academia, and especially science and technology, are male sex-typed professional fields (Acker, 1990; Gilbert, 2007). Moreover, the prototypical manager as well as the ingenious scientist are stereotypically not only individuals with masculine traits, but are envisioned as male (Döge, 2006; Singh et al., 2006). When functions and hierarchical positions seemingly require stereotypically masculine traits and qualities, it is expected these traits and qualities are ascribed to men more easily than to women (Schein, 2001). Concerning job competition, this might provide men with an advantage over equally qualified and suited women in the assessment of their talent. This situation happens potentially via implicit, pre-conscious associations of relevant qualities with individual men and women.

Empirically, the phenomenon of gendered assessment of equalized performance by men and women in a male-dominated field has been studied for rather limited contexts. An early laboratory experiment examined readers' evaluations of academic output in the form of articles manipulated to be authored by men and women. Subjects rated women's articles significantly lower than men's for areas of expertise

stereotypically ascribed to men (Goldberg, 1968). The original experiment is methodologically controversial. Replications yielded inconsistent results. A quantitative review of 123 experimental studies designed after the Goldberg experiment found an overall "negligible" gender effect in the assessment of women's and men's achievements. There were high methodological differences in experimental designs, and high variability in effect sizes. Gender bias in the assessment of women's achievements appeared particularly pronounced for job application experiments. It was therefore suggested that the process of selecting individual candidates for job vacancies should be studied more comprehensively in regard to potentially "unique" characteristics and dynamics (Swim, Borgida, Maruyama & Myers, 1989). That suggestion is considered in this work.

A meta-analysis of 61 experimental studies that held all simulated leader characteristics constant excluding gender revealed small, but significant differences in the evaluation of leader effectiveness to the disadvantage of female leaders. Certain context conditions showed higher bias in the assessment of women's competence than others. Evaluation differences became more pronounced when women used leadership styles that represented behavior perceived as stereotypically masculine (Eagly, Makhijani & Klonsky, 1992). It was concluded from these experiments that women are at risk of having their evaluation as effective managers or leaders discounted when assessors perceive incongruity between their gender and professional role requirements (Eagly & Karau, 2002; Eagly, Karau & Makhijani, 1995; Schein, 2001).

Considering the potential impact of (implicit) stereotypes on the assessment of qualified women and men, qualified women and men are expected to be sorted into gendered development tracks in male sex-typed professions. Assessment of past merit and future potential is expected to translate into gendered allocation of rewards and posts. Via allocation of career opportunities based on gender-biased assessment practices, personnel decision-makers actually determine employees' careers within the organization. Empirical results from the field support the hypothesis of a gendered sorting mechanism independent of individual talent: Peer review for post-doctoral fellowship applications in medical science in Sweden was ex post subject to independent re-assessment of applicant scientific excellence by measures of publication quality and productivity. The external review of applicants' scientific merit in comparison with actual selection behavior demonstrated that female scientists were wrongly assessed in their merit and suitability for fellowships by the actual decision-makers in the field. Applicant gender, among other factors, predicted high competence scores in selection practice to the advantage of male applicants (Wennerås & Wold, 1997). Another field study demonstrated gendered assignment of male applicants to high-status jobs and female applicants to lower-status labor queues in recruitment practice despite equal qualification of applicants (Fernandez & Mors, 2008). In organizations with existing merit-based pay practices women and male minority members achieved less pay than white men despite equal scores in performance evaluations. Organizational reward practices which appear meritocratic in nature can therefore produce unequal career outcomes in the form of remuneration for inter alia women (Castilla, 2008). Another field study on recruitment found that women engineers, with qualifications equal to their male peers, are less likely to receive job offers (Shantz et al., 2011).

These results from the field suggest that women are at a particular disadvantage when competing with men over male sex-typed jobs, whether they are high-status positions or functionally located in a stereotypically masculine field of expertise. In a pronouncedly male sex-typed organizational context, selection and promotion practices are expected to create differential career outcomes for qualified men and women. Consequently, we hypothesize that

H1: the visibility of applicant gender results in a more positive assessment of men and a more negative assessment of women's suitability and qualification for a male sex-typed job vacancy.

and furthermore,

H2: the visibility of applicant gender allows decision-makers to more strongly perceive and stress stereotypically masculine and feminine qualities in male and female applicants.

2.2 Homophily and Team "Fit" as Sources of Biased Individual Assessment

Considering an organizational context that is highly demographically homogeneous, majority status, in salient attributes or features, might play a role in personnel selection processes. Perceived similarity may spark positive associations, feelings of cohesion and trustworthiness, shared values and comfort. Homophily may cause individuals to ascribe favorable attributes to peers they perceive as "the same". Theoretical contributions have suggested that there could be a relation between perceived team fit and performance expectations. Fit-derived performance expectations could potentially cause favorably flawed evaluation processes for performance expectations (Dipboye, 1985; Heilman, 1997). This situation could lead to positively flawed organizational selection and recruitment practices to the benefit of individuals fitting the current composition of the organizational workforce (Burke, 1994; Granleese, 2004). Thus, perceived team fit may constitute a decisive factor not only for positive evaluations, but also for preferred selection. With regard to organizational context, gendered selection preferences may also serve a perceived need to establish compatibility between current organizational members, culture and stakeholders of the organization (Fawcett & Pringle, 2000). Within homogeneous teams – in particular the upper echelon of organizations – several authors have suggested that in-group member status may constitute an advantage in being selected to serve on boards of directors (Rindfleish & Sheridan, 2003; Terjesen, Sealy & Singh, 2009; Van der Walt & Ingley, 2003).

Observing demographic similarity to current organizational members and decision-makers might theoretically be a factor leading to privileged career advancement. Empirical examination of such phenomena is necessary. Underlying motivation of the selection and promotion of similar peers might be – in line with the described approaches of the social sciences – associations of trust with social similarity. We therefore hypothesize that in a highly male-dominated field, decision-makers are expected to more readily entrust male applicants with particularly important and

challenging assignments. When "team fit" is required, such fit will be established through gender similarity, where gender is a salient feature.

H3: A necessity of "team fit" will work to the disadvantage of qualified female candidates in evaluation and selection outcomes.

2.3 Effectiveness of Anti-Discrimination Law as Behavioral Moderator

Organizational recruitment practices happen in the regulated area of employment law. Unfavorable employment outcomes for members of certain legally protected social groups, like women, carry legal implications for employing organizations. Organizational efforts to comply with equal opportunity laws have not yet been considered as a factor in studies on unequal employment outcomes for women. Incentives for organizational compliance should in theory ensure the existence of effective recruitment policy measures to counter individual or institutionalized bias in applicant evaluation, selection and treatment in Europe (Masselot, 2007) and other countries that afford legal protection against gender discrimination in employment. In addition to laws, equal opportunity is one of the most frequently voiced organizational values (Kaptein, 2004). While no empirical studies have yet tested for organizational legal compliance efforts in the personnel selection process, there is a laboratory experiment on recruiter discrimination against ethnic minority members concerning codes of conduct. When decision-makers were asked to consider team cohesion in their selection, the existence of codes of conduct – with an explicit equal opportunity affirmation – was per se not able to prevent discrimination against foreign applicants. Codes of conduct with the threat of sanctions were able to counteract such discriminatory selection practice (Petersen & Krings, 2009). When comparing organizational members' obligations to comply with behavioral instructions arising from codes of conduct or laws, responsibilities to comply with laws generally rank higher than "solely" ethical responsibilities (Carroll, 1979) embodied in codes. A principally fundamental importance of legal norms for organizational members may, however, not enter into effect for the area of anti-discrimination law. It has been suggested that these provisions may suffer from low effectiveness since it is "difficult" to legislate against decision-maker bias (King et al., 2011). Viewing gender stereotypes and implicit associations might affect decision-makers on a pre-conscious level (Kiefer & Sekaquaptewa, 2007), equal opportunity laws might fail to effectively tackle "modern" – covert and subtle – forms of gender discrimination. An organizational practice that requires compliance with anti-discrimination law with no threat of sanctions is therefore not expected to constitute an effective guideline for individual behavior. We therefore hypothesize that

H4: instructions with no enforcement or accountability threat will not be able to counteract existing (implicit) gender bias to the disadvantage of qualified women.

3 Method

Striving to complement existing empirical research on gender discrimination in recruitment, hypotheses derived from theory were tested in the controlled context of the laboratory. Some factors were held constant while others could be varied so that the impact of stereotypes and implicit associations, homophily and team fit, and instructions to respect anti-discrimination law in recruitment were to be examined in a recruitment simulation. The experiment was to be set in a male-dominated field, and science and technology was chosen for this purpose.

The experiment was conducted at the Vienna University of Technology, henceforth abbreviated “University”, which is respected for its achievements in education and research in SET. The University had become aware of unbalanced recruitment and retention rates of women in comparison to the total talent pool available on the junior scientist level. University leadership wished to identify reasons for its disproportionate loss of women along the hierarchy. Three distinct schools were selected for participation in order to grasp the diversity of disciplines within SET. The selected contrasting schools are Mechanical Engineering, Civil Engineering and Technical Chemistry. In 2011, they accommodated 10%, 21% and 36% women as students, respectively (Koordinationsstelle für Frauenförderung und Gender Studies, TU Wien).

3.1 Experiment Design

A single-blind laboratory experiment was designed after Petersen and Dietz (2000) and Petersen and Kring's (2009) experimental work: These model experiments created a controlled context for the observation of discriminatory decision-maker selection behavior. They observed discrimination on the basis of origin under cues to establish team cohesion by superiors. They also studied the impact of a normative framework in the form of a code of conduct requiring equal opportunity independent of origin. This work is built upon and complemented in line with the proposed hypotheses on implicit gender stereotypes, homophily in the male-dominated context and compliance requirements in form of the law.

Personnel selection required a context where both men and women would realistically apply, which tends to be the case for junior positions. Thus, the experiment simulated a recruiting process of university graduates for a junior research position.

The design needed to account for potential social desirability bias as a well-known challenge to laboratory experiments on controversial or shunned human behaviors like discrimination (Powell, 1987; Roth, Huffcutt & Bobko, 1998). If participants suspected their behavior was under scrutiny, they would depict socially-welcomed rather than realistic behaviors (Webster & Sell, 2007). The threat was exacerbated as organizational members knew university leadership had commanded a study on potential gender bias. Subjects were informed that the current recruitment practice of representatives of the technological scientists at the University would be studied in the experiment; their priorities and thoughts on fictional candidates were of interest. A strategy to manage social desirability bias was to change the experimental task from

the actual institutional context by having subjects assume the fictitious identity of a senior scientist at a prestigious private research institute. This change from subjects' own institution was expected to create a form of "disembedment" from the familiar context (Güth & Kliemt, 2010), which may liberate participants from social pressures in their responses.

Targeted subjects for the personnel selection experiment were full professors, senior scientists and advanced students of the three schools Mechanical Engineering, Civil Engineering and Technical Chemistry. The tasks were adapted to discipline-specific niches for all three faculties in order to realistically simulate personnel selection and to create an equidistance for all senior scientists to the task in question.

The experimental task required a specialized senior scientist from a research institute to support HR in personnel selection for a junior position. The junior position was to be filled in a highly prestigious, industry-funded research project. The task was to (a) rank three out of six principally qualified candidates for a job interview, and to (b) justify their choice in writing. Participants were randomly allocated to one of four experimental treatments (see Table 1):

The first treatment, the basic neutral scenario (i.e. control group), presented six complete curriculum vitae in which any information regarding gender had been removed. Candidate curriculum vitae manipulations were limited to two fundamental levels: the qualification level was "excellent"/"good" and the presentation of motivation letters was stereotypical "masculine"/"feminine". The latter manipulation manifested itself in the motivational letters' wording and personal interests (Steffens & Mehl, 2003). Stereotypical "masculinity" expressed itself inter alia in assertive self-presentation in the motivation letter and in "tough", mono-disciplinary research interests in which men on average more readily engage. Stereotypical "femininity" expressed itself through a focus on team rather than individual achievements and an interest in interdisciplinary research. Relevant research topics were manipulated with experts in the discipline inside and outside of the University. Excellence in the research area was manipulated through outstanding academic achievement and a diploma thesis experience that was directly relevant and applicable to the research project in question. Good qualification was characterized by slightly less outstanding academic achievement and relevant, if not directly applicable, research experience. In a conducted pilot, sound manipulation of excellent vs. good qualification was identified as a major challenge. The manipulation of qualification levels with experts in the respective fields inside and outside the University was given maximum attention. Exactly three out of six applicants were manipulated to have excellent qualifications. Since ex aequo rankings were not permitted, the task had an unambiguous solution.

On the excellent qualification plane, a perfect twin pair of stereotypically "masculine" applicants is manipulated to differ only by their gender in gender-visible experimental treatments ("Eva" and "Leo"). This is mirrored on the good qualification level with another perfect twin pair ("Zoe" and "Ben"). For both of these pairs, the applicants' alma mater was a randomized University/non-University (internal/external education from the participant's real-life institution), in order to control for biases stemming from the reputation of the universities. Diploma theses' topics and titles were randomized between the perfect pairs in order to exclude undesired diploma thesis preferences in candidate evaluation and ranking.

Table 1
CV Manipulations across Experimental Treatments by Applicant

	AMY ♀	EVA ♀	LEO ♂	BEN ♂	ZOE ♀	JAN ♂
CV characteristics	feminine	masculine	masculine	masculine	masculine	feminine
Qualification	Excellent	Excellent	Excellent	Good	Good	Good
Alma mater	internal	Randomized int/ext	Randomized int/ext	Randomized int/ext	Randomized int/ext	external

To increase task complexity, and thereby render it more difficult for respondents to figure out the design's underlying research questions, a third candidate on both excellent and good levels was introduced. The third candidate was stereotypically "feminine" and a solely internal ("Amy") or solely external applicant ("Jan"). Complexity of the task was consciously used to further counteract social desirability bias. When not only gender, but alma mater (insider/outsider status) and stereotypical "masculinity" and "femininity" were potentially perceived as relevant to the task by subjects, the chances that they would strive to respond in a perceivably socially desirable way would be decreased. Amy as an excellent, stereotypically feminine applicant was manipulated as an inside candidate across the entire sample, whereas both twin pairs were randomized to be inside or outside applicants. Through this insider manipulation, Amy was created positively unrivalled. She was expected to function as an outlet for subjects' social desirability bias. Once subjects would have felt they had made the "desired" choice by ranking Amy first, potential discrimination of equally qualified and oriented applicants could still be observed in the treatment of the perfectly identical twins.

The second treatment or basic visible scenario added names, randomized photos and gendered information such as military service to an otherwise identical curriculum vitae from the neutral scenario. Using the basic neutral scenario as a control scenario, hypotheses 1 and 2 could be tested using this treatment.

The third treatment added a homophily/team fit cue in the form of a subtle incentive to favor male applicants in order to test hypothesis 3. The HR officer, who asked for expert support in the assessment of candidates, suggested giving thought to team fit and expressed a subliminal preference for stereotypically "masculine" qualities like assertiveness, to ensure the continuation of excellent cooperation from project-funding industrial partners.

To test hypothesis 4, the fourth treatment added a piece of information regarding equal opportunities law, with all other things remaining equal. Institute leadership reminded internal personnel decision-makers that equal opportunity law had to be respected in all recruitment decisions.

3.2 Sample and Data Collection

The full sample contained responses from 296 subjects, which is a large sample for an experiment. While most laboratory experiments draw from student populations this

study counts close to 30% of the highest and high-ranking professionals among respondents (38 full professors and 52 senior scientists).

There was a particularly high response rate of 76% amongst professors. The experimental task was sent to professors via internal mail and included a letter from University leadership asking for their participation. The full sample of senior scientists from two schools – Mechanical Engineering and Civil Engineering – was invited via personal email to participate in an on-site experimental setting. Response was very low despite incentives (4.62% and 5.32%, respectively). Senior scientists from Technical Chemistry were subsequently contacted in the same way as professors, a strategy that substantially increased response. Random samples of advanced students were achieved on-site in suitable diploma/master classes from the respective disciplines.

Table 2
Subjects in Treatment/Faculty/Hierarchy

Sample	Mech. Eng.	Civil Eng.	Chemistry	Total
Men	95	60	79	234
Women	16	17	29	62
Professors	14	14	10	38
Senior scientists	11	9	32	52
Students	86	54	66	206
Treatment 1 (blind)	29	26	31	86
Treatment 2 (regular)	28	21	34	83
Treatment 3 (regular, team fit)	29	15	19	63
Treatment 4 (regular, compliance)	25	15	24	64
Total	111	77	108	296

In total, 21.05% of subjects were women (0% of professors, 19.61% of senior scientists, 25.51% of students). 8.16% of senior scientists and 13.85% of students indicated their nationality was non-Austrian. The average age of senior scientists was 43.27 years. Students were on average 22.21 years old. Professors were not asked to provide demographic detail, since it was feared that this would diminish trust in their anonymity and hence reduce willingness to participate. 29.05% responded to the gender-blind scenario; 70.95% to the three gendered treatments altogether.

4 Results

The experiment delivered two types of data: (1) shortlists of candidates which were analyzed quantitatively by applying paired comparison models and (2) verbal justifications for shortlists which were analyzed using qualitative content analysis. In

the following area, analysis and results of both the ranking patterns and their subsequent justifications are presented.

Analysis of Ranking Patterns

Subjects shortlisted their choice of three from the six applicants in a ranked pattern (from "1", best applicant, to "3", third best applicant), to invite for a job interview. Applicants represented "bundles" of characteristics – social information, professional and academic details and self-descriptions. Ranking of "bundles of characteristics" allows for the attachment of relative importance to appreciated and undesirable traits by subject, and requires cognitive effort visible in reflective judgments, which is expected to create good quality data (Dabic & Hatzinger, 2009, p. 131). A ranking of applicants as complex personalities and "full packages" of skills and characteristics is both realistic and offers methodological advantages compared with ratings. Were subjects asked to rate different applicants, they might be tempted to award artificially inflated rating scores, in order to avoid difficult choices between compared qualities (Krosnick & Alwin, 1988). Ex aequo ranks were prohibited in order to take full advantage of the methodological advantages of rankings. A potential downfall of ranking is that cognitive effort invested may decrease for lower ranks. In this experiment, the maximum number of ranks to request was determined to be three out of six. Threats of less diligent lower ranks were thus avoided (ibid.). As a consequence, only partial rankings were available, which impacted the choice of analytical method.

A Log Likelihood Bradley-Terry's model (LLBT) (Bradley & Terry, 1952), which is the standard approach to modeling paired comparisons, was used (see Appendix for a detailed description of the theoretical foundation of the model).

The parameters were estimated using standard software in the form of the package prefmod for R (Hatzinger, Dittrich & Salzberger, 2009; R Development Core Team, 2011). Data was partially ranked as only three out of six applicants were ranked according to preference for an interview. Partial rankings in that hierarchy of preferences imply that preferences for applicants are not independent. The fundamental assumptions of the estimated model are that applicants ranked are preferred over non-ranked applicants, and comparisons between applicants that are not ranked are missing (MCAR) (Dittrich, Hatzinger & Katzenbeisser, 1998).

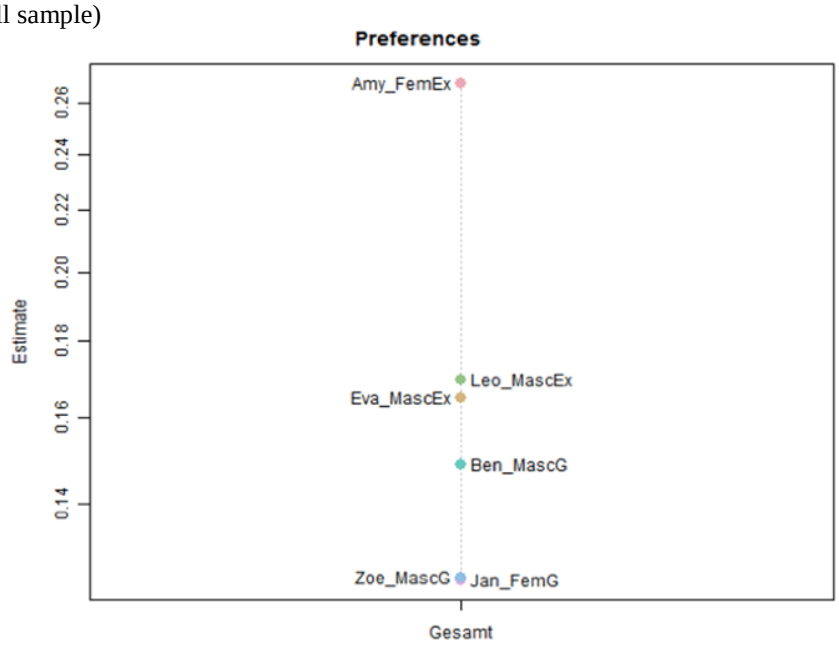
Subject co-variates and interactions of subject co-variates were added to the basic model to determine whether they had an impact on preference orders. The respondent categorical co-variates considered were the experimental treatment allocated to subjects, their gender, age, professional group of respondents and faculty or discipline (Hatzinger, 2011; R Development Core Team, 2011).

The null model established the preference order of the six ranked items for the entire sample. Subject co-variates were not yet considered for distinction. For the full sample and its undistinguished preferences, the applicant ranked last – "Jan"– was used as the reference item for all other candidates ("JanFemG", zero). The null model (LLBT) was hence estimated with 5 object parameters, "AmyFemEx", "Eva_MascEx", "Leo_MascEx", "Ben_MascG", "Zoe_MascG") (Hatzinger et al., 2009, p. 8). The null model yielded the preference order of item parameters across the entire sample (deviance: 1.657,93; log likelihood -1266.31). All item parameter estimates were positive. Overall, positively unrivalled, strictly inside candidate "Amy" scored the highest object parameter, followed by perfect twins "Leo" and

"Eva" within a relatively close distance. As for the "good" candidates, for the overall sample, "Ben" fared considerably better than equally qualified twin "Zoe", whose object parameter estimate was close to zero. For the overall sample, "Zoe", twin to "Ben", was the only candidate that for the complete sample was not significantly preferred over "Jan", the only outside applicant.

Object parameters re-parameterized to worth parameters for the sake of a graphic depiction on a normed preference scale yielded the following preference order = 1):

Figure 1: Null model – Paired comparison patterns depicted on a preference scale (full sample)



Personal traits of experimental participants were considered in order to gain a better understanding of their impact on selection decisions. As model complexity was increased for the basic LLBT through the inclusion of subject co-variate parameters, the following enhanced models are relevant in the sense that some of them provide significant increases in the explanatory power of the model:

Table 3:
Model selection I – Improvements in model fit based on likelihood-ratio tests

Model	Deviance difference to null model	df	p-value
Professional Group	16.8926	10	.0768[#]
Experimental Treatment	32.3368	15	.0058**

[#] *near-significant at p < 0.1; ** significant at <.01*

Gender or age group did not improve model fit. The professional group of subjects – full professors, senior scientists, students – brought a near-significant contribution to the null model ($p < .1$). Experimental treatment contributed significantly ($p < .01$, with a satisfying deviance difference of 32.33 over 15 degrees of freedom).

The subject co-variate experimental treatment enabled the testing of ***hypotheses 1, 3*** and ***4***. The switch from gender-blind – control-treatment "Basic Neutral" – to a gender-visible scenario (treatment 1: "Visible Basic") isolated the effect of gender revelation on selection results. A cue to select applicants based on team fit (treatment 2: "Visible Incentive") tested the impact of homophilous tendencies suggested by important stakeholders. Finally, a cue to respect equal opportunity law (treatment 3: "Visible Law") added the opportunity to observe the effect that a reference to anti-discrimination laws had in recruitment. Table 3 contains object parameters' change compared to the null model across the added experimental treatments:

Table 4
LR test of LLBT treatment type vs. LLBT null model

Candidate	Type	Qualif.	λ_j / λ_j^{OS}	SE	z-value	p-value
Amy (Basic Neutral)	Fem.	Exc.	0.4037	0.0579	6.971	<.001***
Eva (Basic Neutral)	Masc.	Exc.	0.2700	0.0526	5.131	<.001***
Leo (Basic Neutral)	Masc.	Exc.	0.2124	0.0517	4,107	<.001***
Ben (Basic Neutral)	Masc.	Good	0.0688	0.0528	1.303	.193
Zoe (Basic Neutral)	Masc.	Good	0.0196	0.054	0.363	.717
Amy (Visible Basic)	Fem.	Exc.	-0.0152	0.0832	-0.183	.855
Eva (Visible Basic)	Masc.	Exc.	-0.2527	0.0718	-3.522	<.001***
Leo (Visible Basic)	Masc.	Exc.	-0.1271	0.0707	-1.797	.072#
Ben (Visible Basic)	Masc.	Good	-0.0691	0.0722	-0.957	.3386
Zoe (Visible Basic)	Masc.	Good	-0.0801	0.0741	-1.080	.2801
Amy (Vis. Incentive)	Fem.	Exc.	-0.0480	0.0865	-0.554	.5796
Eva (Vis. Incentive)	Masc.	Exc.	-0.1274	0.0791	-1.611	.1072
Leo (Vis. Incentive)	Masc.	Exc.	-0.0283	0.0783	-0.361	.7181
Ben (Vis. Incentive)	Masc.	Good	0.1084	0.0796	1.361	.1735
Zoe (Vis. Incentive)	Masc.	Good	0.0771	0.0805	0.958	.3381
Amy (Visible Law)	Fem.	Exc.	0.0135	0.0895	0.151	.8800
Eva_(Visible Law)	Masc.	Exc.	-0.1349	0.078	-1.730	.0836#
Leo_(Visible Law)	Masc.	Exc.	-0.0552	0.0773	-0.714	.4752
Ben_(Visible Law)3	Masc.	Good	0.0799	0.0785	1.018	.3087
Zoe_(Visible Law	Masc.	Good	-0.0500	0.0824	-0.606	.5445

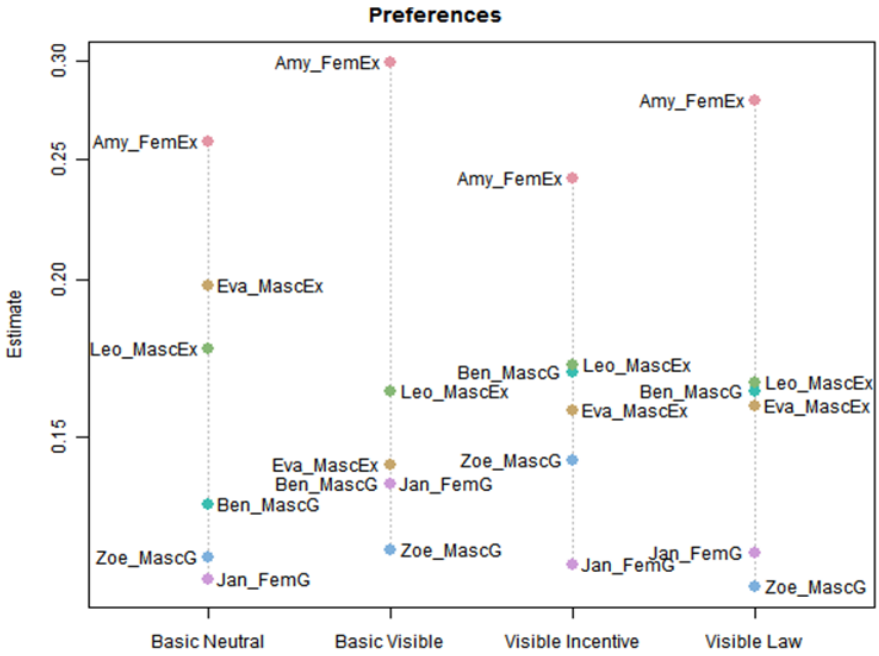
*# near-sign. at p < .1,· * sign. at p < .05, ** sign. at p < .01, *** highly sign. at p < .001*

Once candidates become identified by their gender, the object parameters revealed that *Eva* experiences a pronounced loss in her worth parameter when visible as a

woman compared with the gender-blind control treatment. Excellent *Eva*'s decline with the revelation of her gender was the only significant change for a single applicant by itself across treatments. The finding is highly significant ($p < .001$).

As object parameters were re-parameterized to worth parameters for graphic depiction on a scale, the plot (see Figure 2) shows inside-candidate *Amy* ranked consistently at the top of the applicant lot. The results show clear differences in selection outcomes (illustrated on preference scales) for individual experimental treatments for the perfect pairs *Eva*/*Leo* and *Ben*/*Zoe*.

Figure 2: Paired comparison patterns on preference scales by experimental treatment



Since the subject co-variate treatment's model itself brings highly significant explanatory power, other findings that are non-significant by themselves shall be considered in their entirety, albeit with caution regarding individual candidate differences. *Eva*'s perfect twin *Leo* upon revelation of gender has a slightly lower probability of being selected over all other candidates to which he was compared. *Leo* gains second rank after *Amy*. At the same time, on the good qualification plane, *Zoe* loses a rank to *Jan*, whom she surpasses as long as she is not identified as a woman. Once homophily and team fit are added, *Eva* is not able to regain the second rank she achieves as long as her gender is unknown. On the contrary, she loses in rank not only compared to her twin of equal qualification, but also compared to worse-qualified *Ben*. For *Zoe*, the requirement of homophily shows a slightly positive effect: She gains an upper hand over outside candidate and stereotypically feminine *Jan*. The

latter surpasses her when identified as a man without allusion to stereotypically masculine job requirements.

When compliance with anti-discrimination law is required, the result does not change to the benefit of the two women *Eva* and *Zoe*: *Eva* remains at rank four despite her making part of the top three excellent candidates. The result is that she does not receive an invitation to a job interview behind equally qualified *Leo* and worse qualified *Ben*. *Zoe* finds herself behind both of her equally qualified competitors again, even behind outside applicant *Jan*.

In summary, the quantitative analysis of preference orders reveals that experimental treatment exercised significant influence on the selection behavior of subjects. Our results demonstrate that once gender is visible, this criterion is likely to be singled out as a critical inferential point about how suitable an individual applicant is, as has been previously suggested in the social psychology theory (Heilman, 1997, p. 882). Early literature reviews came to indecisive conclusions about what fate "exceptionally" qualified women faced in the evaluation of their qualification (Swim, et al., 1989). The findings of the present study, however, clearly show that women of excellent qualification are disadvantaged by their gender alone when compared to male competitors in a male-dominated field. They even experience this disadvantage if they dispose of all the right skills and present themselves to be no different than their male competitors in a stereotypically "masculine" way (*Hypothesis 1* confirmed).

It has previously been suggested that gender bias may depend on situational factors (Basow, 1986). For instance, it has been suggested that decision-makers are willing to engage in discrimination if they perceive important others to favor this behavior, as part of a "rational bias" (Porter & Roberts, 1976; Szwajkowski & Larwood, 1991). The results of our experiment support these findings: Excellent women face increased bias and decreased chances of being ranked for an interview once superiors encourage preferential hiring of men by appealing to feelings of team fit and cohesion. Results, therefore, suggest that not only tangible career benefits, but also social aspects like protection of team cohesion may cause personnel decision-makers to place socially different applicants at a disadvantage. (*Hypothesis 3* confirmed).

A situational stimulus to refrain from discrimination for normative reasons, even when issued by highest leadership, does not exercise an effect to mitigate discrimination against women (*Hypothesis 4* confirmed). This result supports earlier findings of Petersen and Kring's (2009) model study, which confirmed that the mere existence of a code of conduct is insufficient in preventing discrimination. The same appears to be true for the law as an extra-organizational body of norms.

The results also point to another important aspect: The effect of gender bias in the assessment affects women of both excellent and good qualification. It is important to note that the excellent female applicant not only faced disadvantage with her "twin", but also with a worse-qualified applicant. This is observed when team fit is relevant in a male-dominated professional context. Acknowledging that team fit seems to be primarily assessed via index features like gender carries implications for other homogeneous, male-dominated teams and selection into them. Hence, our results corroborate theoretical suggestions from the field of gender studies: e.g. Metz (2003) suggests that for top teams in corporate management, tangible qualities like performance might be overpowered by social "fit". Furthermore, it was previously

suggested women might be tested to a higher degree before being admitted to predominantly male corporate boards (Rindfleish & Sheridan, 2003), and women might be at a particular disadvantage when a job vacancy is associated with "tough" or requires "masculine" traits (Fawcett & Pringle, 2000; Heilman, 1997).

Analysis of Ranking Justifications

Written justification of selection arguments yielded rich qualitative data. Professors' justifications for the analysis were used as they are the most experienced decision makers in the sample. Respondents predominantly structured their selection justification into concrete arguments, noting advantages and disadvantages of applicants in check-list or flowing text form. This material was unitized into 544 sense units (Srnka & Koeszegi, 2007) and coded into 121 categories. To ensure reliability, two independent coders coded 31.6 % of the sense units in parallel. A comparison of the coding and the calculation of Cohen's Kappa as a measure of inter-coder reliability revealed extremely satisfactory inter-coder consensus at 91.08 %. 327 of the sense units provided positive or negative arguments for the ranking of individual applicants. 75.8% of these arguments were positive arguments, expressing qualities or advantages of individual applicants. Respondents argued primarily in the affirmative, hence *for* the candidates, and to a far lesser extent *against* applicants in their selection justification. It can thus be presumed that these predominantly positive arguments will contain the key criteria underlying selection reasons.

The justifications can be separated into six main quality categories: General qualification, academic qualification, practical experience, functional-technical, personal skills and language skills. It is hypothesized that men in a male sex-typed and dominated context will be ascribed more qualities – overall quality mentions – than women once gender is made visible (**Hypothesis 2**). This hypothesis is refined after the development of the main categories. Men in a male sex-typed and dominated context will be ascribed more personal qualities than women once gender is made visible (**Hypothesis 2a**). Observing the sex-typed nature of some of the categories, it is expected that function-technical ability will be more readily ascribed to male than female applicants once gender is visible (**Hypothesis 2b**). The contrary holds for language skills as a feminine sex-typed propensity. Women are expected to more readily be ascribed language skills compared with men once gender is revealed (**Hypothesis 2c**).

The following analysis illustrates the differences between the gender-blind control treatment and three gender-visible treatments.[1]

Gender-blind and regular, gender-visible treatments differed in their three most common quality mentions:

[1]Across all four treatments and six main categories, mentions were scarce in some categories. Precision to test the suggested hypotheses would be lost if main categories were to be aggregated in larger – less meaningful – categories for the sake of this analysis. The comparison of gender-blind treatment with the remaining three treatmens where gender was visible appears justifiable viewing salience of gender itself was expected to be a major trigger for gendered stereotypes impacting on assessment of applicant suitability.

Table 5:
Three Most Frequent Positive Applicant Quality Categories

Gender-blind		Gender-visible	
Functional-technical competence	24.14%	Academic competence	34.02%
Personal competence	22.76%	Functional-technical competence	15.98%
Academic competence	21.38%	General qualification	13.92%
	68.28%		63.92%

Functional-technical skills and personal qualities were overall more prominent in relative frequency in a gender-blind context. That personal qualities were also more prominent in a gender-blind context is a finding that might be counter-intuitive considering that name, gender and photo constitute perhaps the most common personal information. In contrast to this, academic competence and general qualification were more frequently a selection argument when applicants were identified as men or women.

Men and women appear to be affected quite differently by the treatments judging from the shift in relative share of quality mentions in gender-blind and gendered scenarios. Table 5 shows changes in applicants' cumulated relative share of competence ascriptions. To simplify comparisons, evaluations for the male and female "twin pair" applications were aggregated. For tests of statistical significance relating to the change in frequency of positive mentions for each category, chi-squared tests could not be applied since expected frequencies were lower than 5 in several cells (Andrés & Tejedor, 1995; Edwards, 1950; Mayo, 1959). Grouping these six main categories into larger categories and utilizing the Chi-square test under Yates' correction would have compromised the opportunity to test the hypotheses. For contexts where frequency of use in responses (Myers, 1958) of two nominal variables in contingency tables is examined and small sample size constitutes a challenge, Fisher's exact test was found to be the most appropriate of the 22 alternative tests (Andrés & Tejedor, 1995; Camilli, 1990).[2] A remaining concern is that Fisher's exact test is considered highly conservative with a rejection rate below nominal significance levels. Fisher suggested departure from strict 5%-levels to take into account this characteristic of the test;[3] this should be considered specifically for small samples. Following this recommendation, significance levels for $p < .1$ will therefore be reported as (near-) significant here in the following discussion.

[2] While some authors consider Fisher and Wilcoxon to be "nearly equivalent", Fisher's exact test holds advantages in power.
[3] This occurs for any discrete statistic contingency tables, but is thought to be compounded by the fact Fisher test conditions on the marginals.

Table 6:
Differences in relative frequency of competence ascriptions[4]

Positive ascriptions	AGGREGATE Leo/Ben			AGGREGATE Eva/Zoe			Fisher test
	Gender Blind	Gender Visible	% change	Gender Blind	Gender Visible	% change	p value
General qualification	2.76%	4.12%	+1.37%	6.21%	3.09%	-3.11%	.962
Academic qualification	4.83%	8.76%	+3.94%	11.03%	11.86%	+0.82%	.889
Practical & relevant experience	2.07%	3.61%	+1.54%	4.83%	4.12%	-0.70%	.895
Functional-technical skills	2.76%	6.19%	+3.43%	15.86%	6.19%	-9.68%	.008**
Personal skills	3.45%	2.58%	-0.87%	9.66%	2.58%	-7.08%	.193
Language skills	0.00%	1.03%	+1.03%	0.69%	1.03%	+0.34%	1.000
Σ/Total	15.86%	26.29%	+10.43%	48.28%	28.87%	-19.41%	.001***

** significant at $p < .01$, *** highly significant at $p < .001$

When contrasting the two aggregate twin pairs, notable changes in competence ascriptions between a gender-blind and a "regular" gendered selection situation occur. Differences arising from the revelation of gender are most pronounced in overall quality ascription and emphasis on functional-technical skills. While the male aggregate collects over 10% more total positive mentions in gendered scenarios compared with the gender-blind evaluation, the female aggregate pair loses close to one fifth of its positive selection arguments (relative to the numbers of positive mentions in the respective scenario in total). The finding is highly significant ($p < 0.001$; *Hypothesis 2* confirmed). The female pair loses close to 10% of functional-technical skill ascriptions, whereas the male pair assumes approximately 3.5% more such mentions with the visibility of applicant gender information. The difference in functional-technical skill ascription to the disadvantage of the female twin pair is significant ($p < .01$; *Hypothesis 2b* confirmed). No significant differences were found between the genders for ascription of personal qualities and language skills. As hypothesized, ascriptions of positive general and academic qualifications and practical experience did not reveal significant gender-specific differences.

Differences in competence ascription were also distinguished by individual applicants. Stereotypically "feminine", inside applicant *Amy* and stereotypically "masculine" *Eva* were contrasted to stereotypically masculine *Leo* (all three have excellent qualifications).

[4] P-values provided in line with hypothesized directions of impact (*hypotheses 2a*, *2b*, *2c*).

Table 7:
*Differences competence ascriptions of **excellent** applicants*

Positive ascriptions	Leo	Eva	Fisher test	Amy	Fisher test
	Change in % with gender revelation		p value	Change in %	p value
General qualification	-0.70%	-2.25%	1.000	+1.71%	.657
Academic qualification	+2.74%	+2.03%	.754	+8.57%	.728
Practical experience	-0.35%	-1.05%	1.000	-1.57%	1.000
Technical competence	+3.78%	-2.61%	.045*	-1.22%	.085#
Personal competence	-2.24%	-3.63%	.889	-5.53%	.894
Language competence	+1.03%	+1.03%	1.000	+3.94%	1.000
Σ/Total	+4.26%	-6.47%	.096#	+5.90%	.418

*# Significant at p < .1; * significant at p < .05.*

With *Leo* as a benchmark for *Eva* and *Amy*, *Eva* loses competence ascriptions by professors in all areas but language skills and academic competence once gender is revealed. All in all, she loses approx. 6.5% in overall quality mentions, whereas Leo gains over 4%. The difference in overall qualities ascribed to *Eva* and *Leo* is near-significant at $p < 0.1$. Excellent *Eva*'s loss in function-technical competence ascription compared with her twin *Leo* is significant ($p < 0.05$; *Hypothesis 2b* partially confirmed). Inside candidate *Amy* gains in competence ascriptions once identified to be a woman; she loses functional-technical skill assessment tested at a near-significant level in comparison with *Leo*'s gain as an identifiable man ($p < 0.1$).

On the good qualification plane, the following dynamics unfold when the gender of twins *Zoe* and *Ben* and stereotypically "feminine", outside applicant *Jan* is revealed to professors:

Table 8:
*Differences in relative frequency of competence ascriptions of **good** applicants*

Positive ascriptions	Zoe	Ben	Fisher test	Jan	Fisher test
	Change in % with gender revelation		p value	Change in %	p value
General qualification	-0.86%	+2.06%	.143	+1.55%	.400
Academic competence	-1.21%	+1.20%	.592	-0.69%	1.000
Practical experience	+0.34%	+1.89%	1.000	-0.35%	1.000
Functional competence	-7.07%	-0.35%	.136	-0.69%	1.000
Personal competence	-3.45%	+1.37%	.024*	+4.12%	<.001***
Language competence	-0.69%	0.00%	1.000	-0.86%	1.000
Σ/Total	-12.94%	+6.17%	0.000***	+3.08%	<.001***

** Significant at p <.05, *** highly significant at p <.001.*

Zoe loses almost 13% of her overall positive ascriptions, whereas *Ben* makes relative gains of over 6%, and Jan gains over 3% once their gender is disclosed. These findings are highly significant ($p < 0.001$; *Hypothesis 2* partially confirmed). *Zoe* experiences the most losses in positive mentions for functional-technical skill and personal competence. Her loss in personal competence is significant when compared with her twin *Ben*'s gains ($p < 0.05$), and highly significant contrasted to *Jan* ($p < 0.001$; *Hypothesis 2a* partially confirmed).

Overall, the analysis clearly shows that male applicants collect significantly more total competence ascriptions in comparison to equally qualified female applicants. Women suffer particular disadvantage in the ascription of functional-technical skills and personal qualities when their gender is revealed. Sense units categorized into the functional-technical competence category contain inter alia references to relevant technical qualifications, to concrete subject areas or methods to use on the job. Under the heading of technical competence, individual applicants are assessed in their construction, problem-solving or development capabilities, as well as in their knowledge of technical fundamentals and basics for future performance on the job. In short, the ascription of functional-technical competence by content of the category mentions appears to describe the "apt technician". These implied ascriptions of competence appear to be decisive for respondents whom they invite for an interview. As long as women are not identified as women and are ascribed qualities including functional-technical qualities, they are awarded career opportunities. Differences in implicit quality perceptions and ascriptions likely crucially determine selection decisions in a perceivably "masculine" and male-dominated field. Therefore, the visibility of gender in recruitment is expected to exert detrimental effects on qualified women's career opportunities in the described organizational and professional context. It is crucial to note here that direct competitors in the form of the "twin" pairs

were manipulated to be of equal professional qualification. Furthermore, women were presented stereotypically “masculine”, with perceivably “masculine” academic and private interests and assertive and independent mindsets and self-presentation styles. For qualified women, doing “all the right things” (having the “right” education, the “right” experience, skills, traits and personality) does not seem to suffice.

Even excellent female applicants suffered observable disadvantage in the ascription of core qualities, particularly of functional-technical skills. For good, stereotypically masculine women, there was a greater negative effect. The disadvantage for women of less than excellent qualification in recruitment may thus be even more severe. This finding runs contrary to early research on gendered assessment of qualification. An early literature review found that when merits of male and female applicants were low, women with low qualifications received more favorable judgments by superiors of their suitability than their male peers (Nieva & Gutek, 1980). If women of excellent qualification already face gendered bias in their qualification assessment, and women of only good qualification face an even stronger bias, it is certainly not an encouraging sign for future change in opportunities for qualified women.

The loss of ascriptions in “masculine” domains was not compensated by an increase in “feminine” quality mentions for the female applicants. The reason for this dynamic could be that implicit associations of “femininity” were never formed; after all, the women in question were manipulated to be stereotypically “masculine”. Another explanation could be that respondents, professors of technological science, did not value stereotypically “feminine” qualities enough to make them relevant selection criteria for vacancies in their fields. Whatever the reason for this behavior, qualified women experienced only disadvantage and no corresponding advantage by implicit evaluations associated with their gender.

5 Conclusions and Limitations

Empirical results of this study confirm that gender, once being discernable in a personnel selection process, is likely to be singled out as a critical inferential point about individual eligibility (Heilman, 1997). This finding implies that organizations willing to attract and retain the best available talent or wishing to improve gender diversity in a “masculine” and male-dominated context should not trust that personnel selection is unbiased. An immediate solution for organizational practice is blind selection, and it should be implemented during the pre-selection of candidates for an interview. Such a step enhances the likelihood that the initial decision will be based on talent and suitability rather than on social filters. Selection in this laboratory experiment proved conforming to the experiment's design – the three excellent applicants were invited for the three interview slots so long as gender was concealed. It can be expected that rational decision-makers who do not consciously discriminate against women will form a relatively solid opinion about an applicant’s "objective” qualification in a blind review. Subjective impressions and biases in an interview will still have an impact. The requirement to provide written and supported assessments of quality, and the necessity to discuss applicant qualification with a second person

beforehand, should help recruiters in their relatively unbiased "blind" assessment. Blind pre-selection will allow women a more fair opportunity to make it to the interview stage, but also potentially benefit other marginalized groups in the same manner.

The inclusion of women, socialized in a male majority setting, does not guarantee rapid change. Rather, raising awareness and implementing effective policies to counter implicit biases are indispensable measures to ensure that qualified people are awarded the career opportunities that they truly merit.

Finally, reminders of the law and organizational policy do not appear to moderate recruiter behavior for the benefit of female applicants. Either the law or organizational policy will need to provide dissuasive sanctions for non-compliance on an individual decision-maker level.

While a carefully designed laboratory experiment has several advantages, the sample composition of our study brings with it limitations as it is tilted towards student participants and male subjects. The proportion of senior decision-makers (close to 30%) is high for comparable empirical projects, and gender ratio in the relevant disciplines is proportionate. Definitive answers on women's selection behavior as compared to men's in male-dominated professions will require further research to be able to infer whether, more broadly, minority individuals might conform to majority notions of “fit” in a given organizational context, despite a perceived human desire to surround oneself with "one’s own kind“. Different surveying methods had to be used for professors and senior scientists in order to create satisfactory participation of high and highest-level decision makers. Several scientific disciplines at the University were studied. While there is reason to believe results may be generalizable to other male-dominated fields, research using similar methodology will be beneficial to understand the described dynamics in related fields and professions.

References

Acker, J. (1990). Hierarchies, jobs, bodies: A theory of gendered organizations. *Gender & Society*, 4(2), 139-158.

Acker, J. (2006). Inequality regimes: Gender, class, and race in organizations. *Gender & Society*, 20(4), 441-464.

Andrés, A. M., & Tejedor, I. H. (1995). Is Fisher's exact test very conservative? *Computational Statistics & Data Analysis*, 19(5), 579-591.

Basow, S. A. (1986). *Gender stereotypes: Traditions and alternatives*. Monterey, California: Brooks/Cole Pub. Co.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324-345.

Burke, R. J. (1994). Women on corporate boards of directors: Views of Canadian chief executive officers. *Women in Management Review*, 9(5), 3-10.

Camilli, G (1990). The test of homogeneity for 2×2 contingency tables: A review of and some personal opinions on the controversy. *Psychological Bulletin*, 108(1): 135-145.

Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. *The Academy of Management Review*, 4(4), 497-505.

Castilla, E. J. (2008). Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, 113(6), 1479-1526.

Dabic, M., & Hatzinger, R. (2009). Zielgruppenadäquate Abläufe in Konfigurationssystemen – eine empirische Studie im Automobilmarkt: Das Paarvergleichs-Pattern-Modell für Partial Rankings. In Hatzinger, R., Dittrich, R., & Salzberger, T. (Eds.), *Präferenzanalyse mit R - Anwendungen aus Marketing* (pp. 119-150). Wien: Facultas.

Dipboye, R. L. (1985). Some neglected variables in research on discrimination in appraisals. *Academy of Management Review*, 10(1), 116-127.

Dittrich, R., Francis, B., Hatzinger, R., & Katzenbeisser, W. (2007). A paired comparison approach for the analysis of sets of Likert-scale responses. *Statistical Modelling*, 7(1), 3-28.

Dittrich, R., Hatzinger, R., & Katzenbeisser, W. (1998). Modelling the effect of subject-specific covariates in paired comparison studies with application to university rankings. *Applied Statistics*, 47(4), 511-525.

Döge, P. (2006). Vom Entweder-Oder zum Sowohl-Als-Auch. Wissenschafts- und Technikkulturen jenseits der Geschlechterpolarität. In: Dudeck, A., & Jansen-Schulz, B. H. (Eds.), *Hochschuldidaktik und Fachkulturen. Gender als didaktisches Prinzip* (pp. 47-55). Bielefeld: UVW Universitätsverlag Webler.

Eagly, A. H., & Johannesen-Schmidt, M. C. (2001). The leadership styles of women and men. *Journal of Social Issues*, 57(4), 781-797.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573-598.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, 117(1), 125-145.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, 111(1), 3-22.

Eagly, A. H., Wood, W., & Diekman, A. (2000). Social role theory of sex differences and similarities: A current appraisal. In Eckes, T., & Trautner, H. M. (Eds.), *The developmental social psychology of gender* (pp. 13-174). Mahwah, NJ: Erlbaum.

Edwards, A. L. (1950). On "The use and misuse of the chi-square test" – the case of the 2x2 contingency table. *Psychological Bulletin*, 47(4), 341-346.

England, P. (1992). *Comparable worth: Theories and evidence*. Hawthorne, NY: Aldine de Gruyter.

Fawcett, R., & Pringle, J. K. (2000). Women CEOs in New Zealand: Where are you? *Women in Management Review*, 15(5/6), 253-260.

Fernandez, R. M., & Mors, M. L. (2008). Competing for jobs: Labor queues and gender sorting in the hiring process. *Social Science Research*, 37(4), 1061-1080.

Gilbert, A. F. (2007). Academic careers in engineering science: Gender effects of recent developments. In *Women in Engineering and Technology Research Paris*, Paris: PROMETEA.

Goldberg, P. (1968). Are women prejudiced against women? *Transaction*, 5, 316-322.

Granleese, J. (2004). Occupational pressures in banking: Gender differences. *Women in Management Review*, 19(4), 219-225.

Güth, W., & Kliemt, H. (2010). Comments on Vernon Smith's-"Theory and experiment: What are the questions?" *Journal of Economic Behavior & Organization*, 73(1), 44-48.

Hatzinger, R. (2011). prefmod: Utilities to fit paired comparison models for preferences. R package version 0.8-27.

Hatzinger, R., Dittrich, R., & Salzberger, T. (2009). *Präferenzanalyse mit R - Anwendungen aus Marketing*, Wien: Facultas.

Heilman, M. E. (1997). Sex discrimination and the affirmative action remedy: The role of sex stereotypes. *Journal of Business Ethics*, 16(9), 877-889.

Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology*, 74(6), 935-942.

Kaptein, M. (2004). Business codes of multinational firms: What do they say? *Journal of Business Ethics*, 50(1), 13-31

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science*, 18(1), 13-18.

King, E. B., Dunleavy, D. G., Dunleavy, E. M., Jaffer, S., Morgan, W. B., Elder, K., & Graebner, R. (2011). Discrimination in the 21st century: Are science and the law aligned? *Psychology, Public Policy, and Law*, 17(1), 54-75.

Kolpin, V. W., & Singell, L. D. (1996). The gender composition and scholarly performance of economics department: A test for employment discrimination. *Industrial and Labor Relations Review*, 49(3), 408-423.

Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *The Public Opinion Quarterly*, 52(4), 526-538.

Laabs, J. L. (1993). The sticky floor beneath the glass ceiling. *Personnel Journal*, 72(5), 35–39.

Masselot, A. (2007). The state of gender equality law in the European Union. *European Law Journal* 13(2), 152–168.

Mayrhofer, M., Meyer, M., & Steyrer, J. (Eds.). (2007). *Macht? Erfolg? Reich? Glücklich? Einflussfaktoren auf Karrieren* (pp. 211-242). Wien: Linde.

Mayo, S. T. (1959). Toward strengthening the contingency table as a statistical method. *Psychological Bulletin*, 56(6), 461-470.

Metz, I. (2003). Individual, interpersonal, and organisational links to women's advancement in management in banks. *Women in Management Review*, 18(5), 236-251.

Myers, J. L. (1958). Exact probability treatments of factorial designs. *Psychological Bulletin*, 55(1), 59-61.

Nelson, R. L., & William, P. B. (1999). *Legalizing gender inequality: Courts, markets, and unequal pay for women in america*. New York, NY: Cambridge University Press

Nieva, V. F., & Gutek, B. A. (1980). Sex effects on evaluation. *The Academy of Management Review*, 5(2), 267-276.

Oakley, J. G. (2000). Gender-based barriers to senior management positions: Understanding the scarcity of female CEOs. *Journal of Business Ethics*, 27(4), 321-334.

Olsson, S., & Pringle, J. K. (2004). Women executives: Public and private sectors as sites of advancement? *Women in Management Review*, 19(1/2), 29-39.

Petersen, L. E., & Dietz, J. (2000). Social discrimination in a personnel selection context: The effects of an authority's instruction to discriminate and followers' authoritarianism. *Journal of Applied Social Psychology*, 30(1), 206-220.

Petersen, L. E., & Krings, F. (2009). Are ethical codes of conduct toothless tigers for dealing with employment discrimination? *Journal of Business Ethics*, 85(4), 501-514.

Petersen, T., & Morgan, L. (1995). Separate and unequal: Occupation-establishment sex segregation and the gender wage gap. *American Journal of Sociology*, 101(2), 329–65.

Petersen, T., & Saporta, I. (2004). The opportunity structure for discrimination. *American Journal of Sociology*, 109(4), 852-901.

Porter, L. W., & Roberts, K. H. (1976). Communication in organizations. In Dunnette, M. D. (Eds.), *Handbook of industrial and organizational psychology* (pp. 1553-1589). New Jersey, NJ: John Wiley and Sons Inc.

Powell, G. N. (1987). The effects of sex and gender on recruitment. *The Academy of Management Journal*, 12(4), 731-743.

R Development Core Team. (2011). A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.

Reskin, B. F. (1993). Sex segregation in the workplace. *Annual Review of Sociology*, 19, 241–270.

Reskin, B. F., & McBrier, D. B. (2000). Why not ascription? Organizations' employment of male and female managers. *American Sociological Review*, 65(2), 210-33.

Rindfleish, J., & Sheridan, A. (2003). No change from within: Senior women managers' response to gendered organizational structures. *Women in Management Review*, 18(5/6), 299-310.

Roth, P. L., Huffcutt, A., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology*, 88(4), 694-706.

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, 57(4), 675-688.

Scott, K. A., & Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organizational Behavior and Human Decision Processes*, 101(2), 230-242.

Shantz, A., Wright, K., & Latham, G. (2011). Networking with boundary spanners: A quasi-case study on why women are less likely to be offered an engineering role. *Equality, Diversity and Inclusion: An International Journal*, 30(3), 217-232.

Singh, V., Vinnicombe, S., & James, K. (2006). Constructing a professional identity: how young female managers use role models. *Women in Management Review*, 21(1), 67-81.

Sosik, J. J., & Godshalk, V. M. (2000). Leadership styles, mentoring functions received, and job-related stress: A conceptual model and preliminary study. *Journal of Organizational Behavior*, 21(4), 365-390.

Srnka, K. J., & Koeszegi, S. T. (2007). From words to numbers - How to transform rich qualitative data into meaningful quantitative results guidelines and exemplary study. *Schmalenbach's Business Review*, 59, 29-57.

Steffens, M. C., & Mehl, B. (2003). Erscheinen „Karrierefrauen“ weniger kompetent als „Karrieremänner“? Geschlechterstereotype und Kompetenzuschreibung. *Zeitschrift für Sozialpsychologie*, 34(3), 173-185.

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989, May). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, 105(3), 409-429.

Szwajkowski, E., & Larwood, L. (1991). Rational decision processes and sex discrimination: Testing ‘rational’ bias theory. *Journal of Organizational Behaviour*, 12(6), 507–527.

Terjesen, S., Sealy, R., & Singh, V. (2009). Women directors on corporate boards: A review and research agenda. *Corporate Governance*, 17(3), 320-337.

Van der Walt, N., & Ingley, C. (2003). Board configuration: Building better boards. *Corporate Governance*, 3(4), 5-17.

Wood, G. J. (2006). Career advancement in Australian middle managers: A follow-up study. *Women in Management Review*, 21(4), 277-293.

Webster, M. jr., & Sell, J. (Eds.). (2007). Laboratory experiments in the social sciences. *Academic Press*.

Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387, 341-343.

Appendix: Paired comparison patterns for partially ranked data

In a classical paired comparison, subjects would be asked to choose one preferred item out of two choices. For *J* items, number of pairs to compare are $\binom{J}{2}$. The response options for a strict choice between items *i* and *j* can be written as follows:

$$y_{ij} = \begin{cases} 1 \; if \; item \; i \; preferred \; to \; item \; j, (i > j) \\ -1 \; if \; item \; j \; preferred \; to \; item \; i, (j > i) \end{cases} \quad (1)$$

Bradley-Terry's model (Bradley & Terry, 1952) is the standard approach to modeling paired comparisons. For the mentioned response in the single paired comparison between two items *i* and *j*, y_{ij}, the probabilities of item *i* being preferred over item *j* and vice versa is dependent on their respective non-negative worth parameters π_i *and* π_j

$$P\{Y_{ij} = 1 | \pi_i, \pi_j\} = \frac{\pi_i}{\pi_i + \pi_j} \quad \text{and}$$

$$P\{Y_{ij} = -1 | \pi_i, \pi_j\} = \frac{\pi_j}{\pi_i + \pi_j} \quad (2)$$

Using the log-linear formulation of the model, probabilities p_{ij} are given as follows

$$p_{ij} = \frac{\sqrt{\pi_i/\pi_j}}{\sqrt{\pi_i/\pi_j} + \sqrt{\pi_j/\pi_i}} \quad (3)$$

The likelihood function is for *l* th response pattern *y* is

$$L = \prod_l P(y_l)^{n_l} \quad (4)$$

For traditional paired comparisons, respondents' choices can be intransitive. Rankings are different insofar as they do not constitute sequential pairwise comparisons. Rather, they produce implied paired comparisons as early choices necessarily effect dependencies. An object is assigned first rank and hence preferred over all other objects at hand. for *J* objects, O_1 when ranked first is preferred *J – 1*

times, O_2 as second choice then is preferred *J – 2* times, etc. The first choice leaves *J – 1* objects open to ranking for the second place, and so forth. Response patterns involving objects that are ranked high and other not yet ranked objects are hence precluded and need to be removed from paired comparison patterns (structural zeros) of all *L= J!* possible patterns.

For identifiability reasons, worth parameters $\{\pi_i\}$ are constrained to sum to one over *J* items. The model hence allows for a determination of scale values for the ranked items (applicants) on a preference continuum that is not directly observable through the ranking itself (Dittrich et al., 1998). Parameters for items compared in pairs *are.* $\gamma_1, \gamma_2, \ldots, \gamma_j = \ln \pi_j$.

This gives

$$\pi_j = \frac{\exp\{\gamma_j\}}{\sum_i \exp\{\gamma_j\}},\ j = 1,2,\ldots,J. \text{ or}$$

$$\pi_j = \frac{\exp\{2\lambda_j\}}{\sum_j \exp\{2\lambda_j\}},\ j = 1,2,\ldots,J. \quad (5)$$

Parameter restrictions require one parameter – typically the last – to be set zero for identifiability (Dittrich, Francis, Hatzinger & Katzenbeisser, 2007).

The expected number of responses where *i* is preferred over *j*, m_{ij}. with m_{ij} = $n_{(ij)}$ p_{ij}. takes the following form for the model's log-linear representation with object parameters λ ($\lambda_i = {}^1/_2\,\pi_i$). and nuisance parameters α:

$$\ln m_{jk} = \alpha_{(jk)} + \lambda_j - \lambda_k. \text{ and}$$

$$\ln m_{kj} = \alpha_{(jk)} - \lambda_j + \lambda_k. \quad (6)$$

The log-linear formulation, increased in complexity due to the heightened number of nuisance parameters in comparison with the logit formulation of the model, is able to include subject covariates (Dittrich et al., 1998). For the inclusion of categorical subject co-variates in the model – in this case experimental treatment, gender,

professional group – each combination of covariates forms a distinct covariate set. For *K* sets – *1 < K <= N – J! = L* response patterns produce *LK* response patterns under inclusion of the covariates. n_{lk} yields the number of times the *l* th response pattern occurs within a covariate set *k.* The linear predictor is

$$\eta_{lk} = \sum_{i<j} y_{ij;lk}(\lambda_{ik} - \lambda_{jk}) \quad (7)$$

Each parameter λ_{ik} gives the interaction effect of item *i* with covariates. Covariates *A* and *B* – for example gender and age – could lead to the effect $\lambda_{i.A} + \lambda_{i.B} + \lambda_{i.A.B}$.

With a covariate *x*, the linear predictor of the model generalizes to the form

$$\eta_{lk} = \sum_{i<j} y_{ij;lk}(\lambda_i + x_k\beta_i - \lambda_j - x_k\beta_j) \quad (8)$$