

**Feminization of management leads to backlash against agentic applicants:
Lack of social skills, not gender, determines low hireability judgments in a
student sample**

MELANIE C. STEFFENS¹, JANETTE C. SCHULT², & IRENA D. EBERT²

Abstract

Effects of an applicant's sex on the ascription of task competence, social skills, and hireability are of theoretical as well as practical concern. In two influential series of studies with American college students, Rudman and Glick (1999, 2001) showed a backlash effect: When social skills were included in job descriptions ("feminized job descriptions"), female, but not male applicants displaying agentic behaviour were rated low in social skills and considered less hireable than their male counterparts. We tested the generality of this backlash effect in a replication with German students. In four experiments a total of 555 participants listened to telephone interviews with alleged applicants. Female and male applicants were rated very similarly with respect to task competence, social skills, and hireability. Feminizing job descriptions resulted in lower hireability ratings of both female and male agentic applicants. These findings held across different conditions (the manner in which applicants spoke, different business areas, subjective and objective rating scales, applicants' parental status, and amount of information received about applicants). Our findings show that both female and male applicants are punished for overly agentic behaviour if social skills are required in a job description.

Key words: gender stereotypes; role congruity; backlash effect; leadership; malleability of implicit stereotyping

¹ Correspondence should be addressed to Prof. Melanie Steffens, PhD, Department of Psychology, Friedrich Schiller University Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany; email: melanie.steffens@uni-jena.de

² Friedrich Schiller University Jena, Germany

When faced with the task of selecting the best-fitting applicant for a vacant position, personnel selectors with expertise in social cognition may be beset with an uneasy feeling. Salient characteristics, for example an applicant's sex, cannot simply be disregarded when forming an impression of the candidate's strengths and weaknesses. One may try to generate as many objective, directly comparable qualification criteria as possible, but an objective "adding up" of applicant experience and accomplishments is certainly limited. Results of published studies are ambivalent as to gender biases in impression formation. The aim of the present research was to understand how far cultural gender stereotypes influence decision making during applicant selection in a relatively realistic scenario. More concretely, our question was whether the hireability of female applicants for leadership positions is considered lower than that of males, and in how far this depends on the type of job description: We tried to replicate the backlash effect against agentic women that has been reported when job descriptions include social skills (Rudman & Glick, 1999, 2001). In addition, we assessed implicit or explicit stereotypes that associate men with agency and women with communion, and whether the impressions formed of individual applicants are related to raters' stereotypes.

Research on gender stereotypes shows that men are traditionally characterized with goal-oriented, agentic attributes, whereas women are characterized with service-oriented, communal ones (cf. Eagly, 1987). Men are described as independent, task-oriented, and dominant, whereas women are considered to be dependent, people-oriented, and sensitive (e.g., Dubno, 1985; Haslett, Geis, & Carter, 1992). However, these coarse traditional gender stereotypes appear to be decreasing over the past decades (e.g., Diekmann & Eagly, 2000; Rodler, Kirchler, & Hölzl, 2001).

What do gender stereotypes imply for potential gender biases in hiring decisions? A major criterion for personnel selection appears to be the perceived fit of an applicant for a given position (Eagly & Karau, 2002; Heilman, 1983). If the perceived fit is good, successful job performance is expected. Traits attributed to women in general do not seem to fit the requirements of a traditionally male-typed managerial position. A female applicant, although she might have identical qualifications as a male applicant, would thus be less likely to be hired because she is perceived less fitting based on the application of stereotypical knowledge. According to role congruity theory (Eagly & Karau, 2002), including social skills in the job description of a manager should increase the perceived fit of female applicants. Such a "feminization of management" (cf. Rudman & Glick, 1999) can actually be observed internationally, with person attributes like "cooperative, able to work in a team, able to soothe conflicts" included in job descriptions. This feminization of management should thus reduce gender bias against female applicants (for a review of relevant findings, see Eagly & Karau, 2002).

From the perspective of role congruity theory, it is surprising that Rudman and Glick (1999, 2001) showed the opposite pattern of findings in two influential series of studies. Those authors argued that being nice is part of the prescriptive female stereotype, and that a woman displaying agentic behaviour would be perceived as violating this stereotype (also cf. Eagly, Makhijani, & Klonsky, 1992, for evidence that female leaders are devalued if they use a directive leadership style). A feminized job description, in turn, provides a justification for using this stereotype violation against the female applicant in hiring decisions. With a traditional job description, in contrast, the stereotype violation would not influence hiring decisions. Rudman and Glick provided empirical evidence for this ironic "backlash effect" in

several studies in which either a traditional (stressing task competence) or a feminized job description (stressing task competence and social skills) was presented. Next, they showed a videotape of a job interview with one of two female or male job applicants. Given the feminized job description, there was more hiring bias against the female agentic applicant. This bias could be traced to a lower rating of her social skills. In addition, participants possessed gender stereotypes that linked agency to males and communion to females, as measured with an Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998). Participants' implicit, but not explicit stereotyping correlated with rating the agentic female as less socially skilled given a feminized job description (Rudman & Glick, 2001). In other words, strong implicit gender stereotypes predicted the backlash against agentic females.

The studies by Rudman and Glick (1999, 2001) address an important theoretical as well as practical problem, which makes it not surprising that they are widely cited. For instance, in a review article written for law experts, based on the Rudman and Glick studies, Borgida and colleagues (2005) concluded that women are discriminated against in hiring decisions: "When women presented themselves as possessing the agentic qualities deemed necessary for leadership positions, they were viewed as lacking in the feminine 'niceness' presumably required for more feminine manager positions." (p. 623) Due to the importance of the backlash effect, conceptual replications of the study that test the generality of the obtained effect seem to be mandatory. An existing study (Steffens & Mehl, 2003) did not replicate the pattern of findings reported by Rudman and Glick (1999, 2001). In that study, cover letters and resumes of female and male young professionals were presented who were allegedly sought for middle-management positions. With a feminized job description, both female *and* male agentic applicants were rated less hireable than with a traditional job description. This pattern of findings is intuitively plausible because no evidence of agentic applicants' social skills was given. So if social skills are part of the job description and no information on applicants' social skills is available, they appear less hireable, regardless of their gender.

It is unclear why patterns of findings differ between these two sets of studies. Using written materials is a procedure commonly used in the Goldberg paradigm where target sex is manipulated while controlling for all other factors (Goldberg, 1968; Swim, Borgida, Maruyama, & Myers, 1989). In this way, possibly biased perceptions of equivalent male and female behaviour can be investigated. The disadvantage of this paradigm is that no individual impression of each applicant can be formed. It could thus be that different impressions are formed of female versus male agentic applicants only if enough individuating information is presented. Whereas in terms of ecological validity, videos of actors and actresses in a pretend job interview (see Rudman & Glick, 1999, 2001) are adequate materials, this approach is less rigorous with regard to experimental control. It is simply not possible to manipulate the sex of an applicant while holding all other factors constant (cf. Wells & Windschitl, 1999).

Overview of the experiments

The aim of the present research was to test whether the backlash effect against agentic females can be replicated with auditory job interviews (also see, e.g., Juodvalkis, Grefe, Hogue, & Svyantek, 2003). Whereas phone interviews are not the most common form of job interviews, they are often used, for instance, to interview candidates from abroad. Using

phone interviews, we minimize the influence of personal features while still fostering the formation of a more elaborate impression than possible from only a written presentation. We investigated how the sex of an agentic applicant (operationalized via a female or male voice) influences the ascription of task competence, social skills, and hireability, given a traditional or a feminized job description. We compared agentic versus communal applicants (Experiment 1), agentic versus androgynous applicants (Experiment 2); and we manipulated whether they spoke in a resolute, agentic or in a kind, communal manner (Experiment 2). Also, the degree of agency the applicant displayed was varied between Experiment 1 on the one hand and Experiments 2-4 on the other. To anticipate: We did not find any evidence of a gender bias in hireability ratings. We then excluded alternative explanations of our findings by carefully probing for guessing of the hypotheses; by manipulating the business area in which applicants were sought (typically male dominated vs. gender neutral); and by using objective in addition to subjective rating scales (Experiment 3). Finally, we tested whether differences in the perception of female and male applicants would emerge if they had children, and if only a short first impression could be formed of them (Experiment 4).

Experiment 1

In Experiment 1, male and female applicants who made agentic versus communal statements were compared. If we replicated Rudman and Glick's backlash effect (1999, 2001), we would observe a hiring bias against the female agentic applicant only with a feminized job description, because her social skills would be rated lower than those of her male counterpart.

Method

Participants

Participants were 143 students of various majors at a large university in former Western Germany, 67% of them female. Their age ranged from 19 to 49 years ($M = 23$, $SD = 5$). They were invited to take part in a study on "personnel selection". In all experiments, psychology students received course credit for participating, and all participants received candy. In this and the following experiments, men were distributed approximately equally across the cells of the design. Due to their low overall proportion, all data were collapsed across participant sex. We set $\alpha = .05$ for all statistical tests, so with 143 participants, effects of $f = .30$ of any of the between-subjects factors could be detected with a statistical power of $1 - \beta = .95$ (Cohen, 1977).

Materials

Job descriptions were tailored to leadership positions in middle management for young professionals. The traditional job description stressed applicants' leadership or agentic qualities (e.g., goal orientation, strategic thinking). The feminized job description additionally

called for traits related to social skills (e.g., supportive, cooperative). The number of words was held constant between both job descriptions. The positions as well as the company (software producer) were only vaguely described in order to avoid potential confounding effects. Job descriptions were pretested and appeared realistic. All materials are available from the first author upon request.

Applicant statements and applicant ratings were closely modelled after those used by Rudman and Glick (1999), except that they were less self-promoting and dominant in order to appear appropriate to us in a job interview situation. In detail, each of the six questions that the applicants were asked during the interview was shown on a computer screen, for instance, “Do you have a philosophy of life?” or “What is your typical role in a team?”, each followed by a reply presented auditorily in a male versus female voice that had been pretested to appear similarly nice and competent. Applicants’ replies were designed so that they either demonstrated agentic or communal behaviour (henceforth, “text content”). For instance, when asked about roles in a team, part of agentic persons’ reply was “I quickly take the leadership position. ... There are situations where someone needs to take control”, whereas communal persons’ replies included “I try to mediate when conflicts and hard feelings arise”.

Participants made *applicant ratings* on 23 items translated from Rudman and Glick (1999, 2001): Nine trait adjectives were related to task competence (e.g., “able to make quick decisions”), and nine related to social skills (e.g., “friendly”). In addition, participants made five overall judgments, one concerning social skills, one concerning task competence, and three hireability judgments (e.g., “How likely would it be for you to hire the applicant?”). All ratings were made on a five-point scale (1, “very improbable” to 5, “very probable”).

IAT. Implicit gender stereotypes were assessed using the IAT described by Steffens and Mehl (2003). The rationale behind the IAT is that people are able to react fast if a pair of closely associated categories (e.g. male – goal-oriented) requires one reaction, and another pair (e.g. female – social), another reaction (stereotype-congruent task). In contrast, if the categories are paired so that they are not associated (e.g. male – social and female – goal-oriented), reaction time should be relatively slow (stereotype-incongruent task). The difference in reaction times between the congruent task and the incongruent task, the IAT effect, is taken as an indicator of the association between the categories (cf. Banaji, 2001). The stimuli used in the IAT were 20 words: 10 male and female first names, and 10 adjectives that represented agency (e.g., independent, ambitious) and communion (e.g. empathetic, cooperative). In addition to three short practice tasks, participants completed two main tasks of 2×60 reactions each. In the stereotype-congruent task, participants were instructed to press a response key on the left as fast as possible if a male name or a trait related with *goal oriented* appeared in the middle of the screen, and a key on the right if a female name or a trait related with *social* appeared. The stimuli related to the concepts *male*, *female*, *goal oriented*, and *social* appeared in the middle of the screen in a random order with a response-stimulus interval of 200 ms. The four concepts were shown in the top corners of the screen throughout each task. The stereotype-incongruent task was identical to the stereotype-congruent task except that the positions of the concepts *male* and *female* were switched. There was feedback on reaction times and number of errors after each task.

Procedure

After initially reading the traditional or feminized job description on the computer screen, participants were asked to form an impression about the applicant whose replies they were about to listen to. Questions were displayed on the computer screen, and replies of the first applicant were presented auditorily via headphone. In the next phase of the experiment, the traits on which the applicant was to be rated appeared in a random order for each participant. After rating the first applicant, participants had the chance to read the job description again, and a second applicant was presented, followed by the respective ratings. A quarter of the participants listened to an agentic male and then to a communal female applicant, another quarter to a communal male and then an agentic female. The order in which applicants were presented was reversed for the other half of the participants. After a subsequent manipulation check that corroborated thorough processing of the job description in each experiment, the stereotype IAT was administered. Finally, participants were thanked and debriefed.

Results

As an indicator of the effect size, R^2_p is reported (see Cohen, 1977). Ratings of task competence, social skills, and hireability (all Cronbach's $\alpha > .76$) were separately analyzed in 2 (text content) \times 2 (applicant sex) \times 2 (job description) ANOVAs. Preliminary analyses showed that the pattern of results for the applicant ratings does not change if only the ratings of the first presented applicant are analyzed, demonstrating that the findings described below are not based on contrast effects from the first to the second applicant. Also, order of presented applicants had no effects. All data are presented in Table 1.

Analyses of backlash effect in agentic-applicant ratings. The top panel of Figure 1 shows task competence, social skills, and hireability ratings of agentic applicants, separately for a feminized and a traditional job description and for female and male applicants. As can be seen, females and males were rated very similarly. In contrast to Rudman and Glick's (1999, 2001) findings, there were no interactions of applicant sex, job description, and text content on task competence, social skills, or hireability ratings of agentic and communal applicants (all $F_s < 1$). Analogously, if only agentic applicants were considered, there were no interactions of applicant sex and job description (all $F_s < 1$). The most powerful test of the backlash-effect hypothesis is a planned Helmert contrast of hireability ratings of agentic female applicants with a feminized job description, tested against the other three combinations of applicant sex and job description (male applicant, feminized job description; female applicant, traditional job description; male applicant, traditional job description). This contrast was not statistically significant ($F < 1$; difference estimate: $-.22$, 95% confidence interval: $-.59$ to $.16$).

More generally, the analyses revealed no effects of applicant sex on task competence or hireability ratings. The only effect involving applicant sex was that more social skills were generally ascribed to female applicants than to males, $M = 3.40$ and $M = 3.20$, respectively, $F(1,139) = 6.56$, $R^2_p = .05$. An interaction effect [$F(1,139) = 4.55$, $R^2_p = .03$] indicated that social skills of agentic applicants were rated lower when a feminized rather than a traditional job description was presented ($M = 2.45$ and $M = 2.29$, respectively), whereas the social

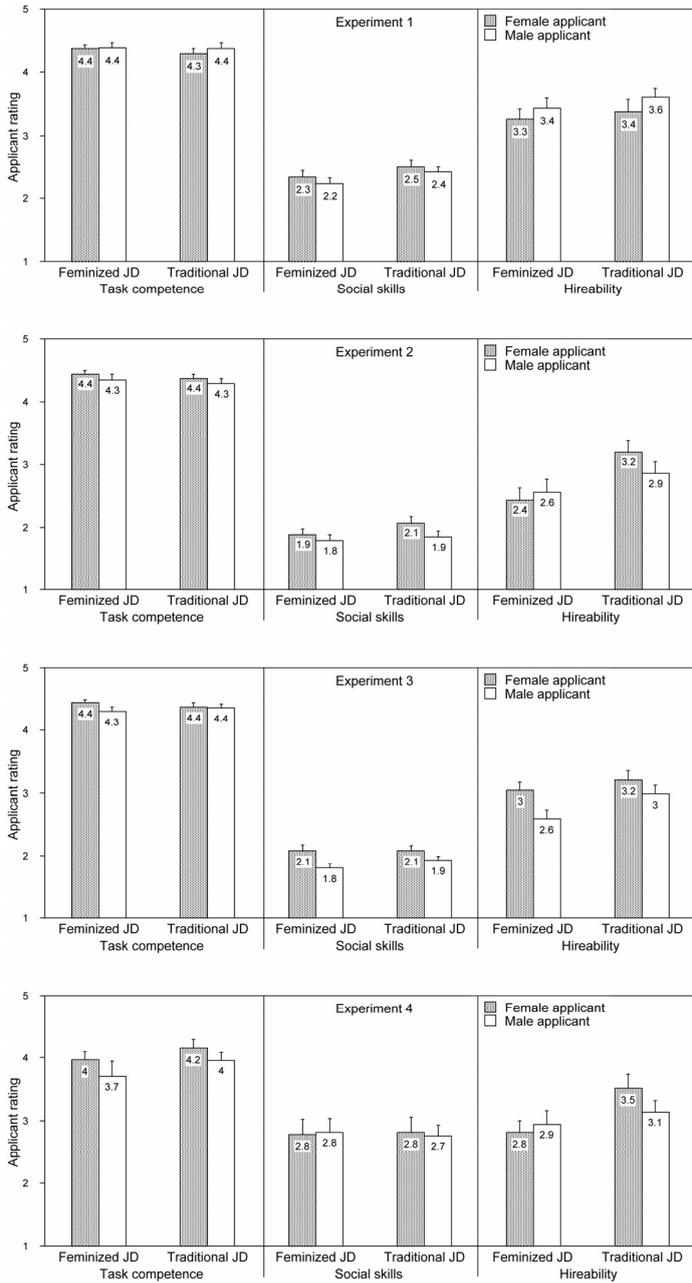


Figure 1:

Mean ratings of agentic applicants with regard to task competence, social skills, and hireability, in Experiments 1-4, separately for a feminized and a traditional job description (JD) and female and male applicants. Error bars show standard errors of means

skills of communal applicants were rated higher when a feminized rather than a traditional job description was presented ($M = 4.15$ and $M = 4.30$, respectively).

Ratings of agentic versus communal applicants. As expected, replicating that part of Rudman and Glick's (1999) findings, we found large text-content effects. In detail, the social skills of communal applicants were rated higher than those of agentic ones, $M = 4.22$ and $M = 2.37$, respectively, $F(1,139) = 628.30$, $R^2_p = .82$. In contrast, agentic applicants received higher task competence ratings than communal ones, $M = 4.38$ and $M = 3.12$, respectively, $F(1,139) = 496.77$, $R^2_p = .78$, and they were rated more hireable than communal ones, $M = 3.42$ and $M = 3.10$, respectively, $F(1,139) = 7.48$, $R^2_p = .05$. There were no other statistically significant main effects or interactions concerning task competence or hireability ratings, but a non-significant interaction of text content with job description that was relevant to the hypothesis, $F(1,139) = 3.07$, $p = .08$, $R^2_p = .02$: Simple main effects tests showed that agentic applicants (both male and female) were considered more hireable than communal ones ($M_s = 3.50$ vs. 2.98) under the traditional job description condition, $F(1,139) = 10.00$, $R^2_p = .07$, but not with a feminized one ($M_s = 3.35$ vs. 3.24), $F < 1$. There were no other effects, all $F_s < 2.13$.

Gender stereotypes. Using traditional IAT scoring (cf. Greenwald et al., 1998), the existence of gender stereotypes was shown through the participants' faster reactions (descriptively: about 90 ms faster) during the stereotype-congruent than the stereotype-incongruent task, $F(1,142) = 133.26$, $R^2_p = .48$. A reliability analysis of the log-transformed IAT effects for the 20 different stimulus words indicated good internal consistency ($\alpha = .86$). Rudman and Glick (2001) had found that with a *feminized job description*, participants with high implicit gender stereotyping rated the social skills of *female* agentic applicants particularly low. However, the relationship we found between IAT scores and social-skills ratings of agentic applicants was not modified by applicant sex or job description (all correlations were negative and numerically larger than $r = -.20$). More concretely, there was a general statistically significant correlation: The stronger the implicit gender stereotypes a given participant demonstrated in the IAT, the lower they rated the social skills of agentic applicants, $r = -.25$. In other words, when there was no information concerning an applicant's social skills (i.e., the applicant displayed agentic behaviour), participants with stronger stereotypes rated (female and male) applicants' social skills lower.

Summary. We did not replicate the backlash effect, according to which only agentic female applicants are rated as less hireable when the job description is feminized. More generally, we found no effects of applicant sex on task competence or hireability ratings. This was the case both with a traditional and a feminized job description. The only factor that influenced these ratings was text content. This finding suggests that applicant sex effects are rather negligible for the ascription of traditional leadership qualities and thus hireability: The way someone presented her- or himself seemed to be more important than whether that person was female or male.

In contrast, the social skills of female applicants were rated higher than those of male applicants, no matter whether the female applicant presented herself as stereotypically communal or counter-stereotypically agentic. In other words, it seems that a gender heuristic was used in social skills judgments. This finding is similar to previous studies, where students and managers gave higher social-skills ratings to female than male agentic applicants on the basis of their resumes (Steffens & Mehl, 2003), and female and male managers ascribed pronounced competence in person-oriented leadership to female managers (Sczesny, 2003).

Given these findings, being female seems to be a sufficient precondition for the ascription of relatively *more* social competence. This stands in contrast to Rudman and Glick's findings (1999, 2001) that agentic females are considered to have lower social skills than agentic males. With regard to implicit gender stereotypes, we also found a different pattern than they did: the more implicit gender stereotypes our participants had, as measured with an IAT, the lower they rated agentic applicants' social skills. Rudman and Glick had reported this correlation only for *female* agentic applicants and a feminized job description.

Thus, we neither found a bias against individual females in task competence nor in social-skills or hireability ratings. However, the interpretation of statistical null effects is always difficult. We can exclude two of three typical problems with the interpretation of a null effect. First, the statistical power of Experiment 1 was comparatively large, and second, the other expected effect that we demonstrated, the enormous effect of text content, shows that our experiment was generally sensitive. A trickier problem is that our manipulation could have missed the sensitive area in which effects arise: It could be, for instance, that presenting oneself as competent and dominant is more forgivable for males than for females only if there is a higher degree of dominance than our applicants displayed (cf. Rudman & Glick, 1999). Therefore, we selected statements that seemed competent and very dominant in Experiments 2-4.

Experiment 2

In Experiment 2 agentic applicants were compared with "androgynous" applicants who portrayed a mixture of agentic and communal traits. Rudman and Glick (2001) had reported that there was no backlash effect against androgynous female applicants. At the same time, androgynous applicants appeared much more hireable than communal ones. Our applicant replies were created to convey task competence in the first two-thirds of each statement. The text content was identical for agentic and androgynous applicants. The last third of each reply conveyed a high degree of dominance for agentic and communion for androgynous applicants.

If we do not replicate the backlash against agentic females, it would still be interesting to obtain further evidence that low social-skills ratings affect hiring decisions with a feminized job description. We therefore manipulated social skills by varying the manner of speaking: One actor and one actress were instructed to speak the respective androgynous or agentic text in a kind, pleasant manner, and another pair of actors spoke in a resolute, very assertive manner. There is a long tradition of research in voice stereotyping. Listeners agree considerably with regard to traits that are inferred from voices, including extraversion, submission (e.g., Allport & Cantril, 1934) and cooperativeness (Addington, 1968). There are cultural stereotypes relating voices to traits (Aronovitch, 1976), and vocal attractiveness leads to halo effects similar to physical attractiveness (Zuckerman & Driver, 1989). More pro-social behaviour is shown in response to a person who speaks with a warm rather than a non-expressive monotonous voice (Goldman & Fordyce, 1983). Even subtle variations like voice volume influence trait ratings, with targets who speak louder being perceived as more aggressive and dominant (Page & Balloun, 1978). In some studies, inter-rater agreement of trait inferences was over $\alpha = .78$, and the traits rated could be subsumed under the dimensions *agency* and *communion* (Berry, 1990, 1991, 1992). Extending these findings, we hy-

pothesized that the manner of speaking influences ratings of the target person's task competence and social skills. We predicted that pleasantly speaking applicants are rated higher on social skills than resolutely speaking ones. In turn, only with a feminized job description, the hireability of resolutely speaking applicants should be judged lower than that of pleasantly speaking ones.

In addition, we investigated the relation between applicant ratings and explicit gender stereotypes, in addition to implicit stereotypes. Therefore, we asked participants to make task-competence and social-skills ratings not only for the individual applicant, but later, to rate whether these traits are, in general, more true of women or of men.

In the final part of the experiment, we sought additional evidence that making dominant statements in a job interview appears more forgivable for male than female applicants, and we tested whether making dominant statements in a job interview appears inappropriate to our participants. Therefore we presented written single statements and participants evaluated applicants and judged whether they would imitate such replies.

Method

Participants

Participants were 192 students of various majors at a large university in former Western Germany (77% females). They ranged in age from 19 to 42 years ($M = 22$, $SD = 4$). Psychology students received course credit for participating. None had participated in Experiment 1.

Materials

The scales on which applicants were rated and the manipulation check were identical to those used in Experiment 1. The traditional and feminized job descriptions were borrowed from a previous study (Steffens & Mehl, 2003). Again, allegedly a middle-management position was to be filled with a young professional, this time in a service company.

Participants listened to three replies of one applicant. The replies were longer than those in Experiment 1 and were changed so that the first two thirds of each reply conveyed agency in an even more assertive, self-confident way than in Experiment 1. Replies of agentic versus androgynous applicants differed only in their last third. Agentic applicants finished with very dominant statements, androgynous ones finished with communal statements. For instance, when talking about roles in a team, the agentic persons' reply finished with "If others talk nonsense, I interrupt them and take over the discussion's leadership", whereas the androgynous persons' reply included "I am happy to help the group reach its aims."

As in Experiment 1, one set of actors were recorded saying these statements as if in a job interview. They spoke in a kind way and will henceforth be referred to as "pleasant-speaking." Another set of actors were instructed to impersonate dominant, stand-offish applicants. They will henceforth be referred to as "resolute-speaking".

Implicit gender stereotypes were assessed with the same IAT used in Experiment 1. Explicit gender stereotypes were assessed by presenting the same trait adjectives on which

participants had rated individual applicants. Each of the 18 traits was rated on a seven-point scale as to whether it was “more true of men” (-3) or “more true of women” (3).

In the final part of the experiment, we presented 9 written single statements, allegedly made by either a male or a female applicant. We asked participants to rate how they liked the applicant and whether they themselves would make such a statement. After pre-studies we selected dominant, self-promoting, and modest statements (three each). Self-promoting and modest statements were constructed following the characterizations of self-promotion and self-effacing in Rudman (1998). Examples are: Dominance: “I am very dominant, so generally, others do what I want.” Self-promotion: “I will successfully master this challenge due to my competencies.” Modesty: “I could be wrong, but I think I might be a suitable applicant for this position.”

Procedure

The procedure resembled that of Experiment 1. After reading a traditional or feminized job description and listening to an applicant’s statements, participants rated the applicant’s task competence, social skills, and hireability. Each participant rated only one applicant who was agentic or androgynous. As in Experiment 1, the manipulation check concerning the job description was then administered, followed by the IAT. Then, explicit gender stereotypes were assessed.

In the final part of the experiment, statements allegedly made by 9 different applicants were presented in a random order, each accompanied by the first and last name of a female or male (e.g., Peter or Petra, counterbalanced), their age, and their study major. Participants rated if they liked each applicant and whether they would make such a statement (each on a 7-point scale). After the first statement, they also indicated whether they remembered that the respective applicant had been male or female.

Results

Applicant ratings on task competence, social skills, and hireability (all $\alpha > .70$) were separately evaluated in 2 (text content) \times 2 (applicant sex) \times 2 (job description) \times 2 (manner of speaking) ANOVAs.

Analyses of backlash effect in agentic-applicant ratings. Replicating Experiment 1, but not Rudman and Glick (1999, 2001), there were no interactions of applicant sex, job description, and text content on ratings of task competence, social skills, or hireability, neither in general (all F s < 1.09) nor specifically for agentic applicants (all F s < 1.41 ; cf. Figure 1). The same Helmert contrasts as in Experiment 1 were computed to compare hireability ratings of agentic applicants. This time, the contrasts were statistically significant, $F(3,91) = 3.03$, $R^2_p = .09$. In detail, with a feminized job description and female applicants, hireability ratings were significantly lower than those in the other three conditions (difference estimate: -.44, 95% confidence interval: -.88 to -.01), a finding that appears to be in line with the backlash effect on first sight. However, hireability ratings for the male applicant with a feminized job description were significantly lower, too, than average ratings of the male and female applicants given a traditional job description (difference estimate: -.47, 95% confidence

interval: $-.94$ to $-.01$). Thus, these contrasts show that the hireability of both female and male agentic applicants is rated lower when the job description is feminized rather than traditional. Moreover, a feminized job description resulted in lower hireability ratings of the agentic, but not the androgynous applicants [interaction: $F(1,176) = 7.20$, $R^2_p = .04$]. There was a simple main effect of job description for the agentic applicants: $F(1,176) = 9.13$, $R^2_p = .05$, $M = 3.02$ and $M = 2.49$, but not for the androgynous ones: $F < 1$, $M = 3.63$ and $M = 3.71$.

In addition, there was a main effect of applicant sex on task competence. Contrary to gender stereotypes, female applicants' task competence was rated a little higher than male applicants', $M = 4.39$ and $M = 4.25$, respectively, $F(1,176) = 6.10$, $R^2_p = .03$.

Ratings of agentic versus androgynous applicants. Confirming expectations, the social skills of androgynous applicants were rated higher compared to agentic applicants, $M = 3.00$ and $M = 1.90$, respectively, $F(1,176) = 55.81$, $R^2_p = .51$, whereas there was no effect of text content on task competence. As to hireability, androgynous applicants were rated more hireable than agentic ones, $M = 3.67$ and $M = 2.75$, respectively, $F(1,176) = 50.64$, $R^2_p = .22$. If social skills were entered as a covariate in the ANOVA, this difference in hireability between agentic and androgynous applicants disappeared ($F < 1$).

Effects of manner of speaking. Ratings of task competence showed, as we expected, that resolute-speaking applicants appeared more task competent than pleasant-speaking applicants, $M = 4.37$ and $M = 4.26$, respectively, $F(1,176) = 4.25$, $R^2_p = .02$ (there were no other effects, all F s < 3.24). In contrast, pleasant-speaking applicants were rated higher in social skills than resolute-speaking ones, $M = 2.61$ and $M = 2.30$, respectively, $F(1,176) = 15.48$, $R^2_p = .08$.

If all agentic applicants rated low on social skills are considered less hireable with a feminized job description, then resolute-speaking applicants should also be rated less hireable with a feminized job description than the other applicants. We tested this hypothesis using a Helmert contrast, testing the combination of "feminized job description, resolute-speaking applicant" against the other three combinations of job description and manner of speaking, in an analysis that also included applicant sex and text content as factors. The contrast variable interacted with text content, $F(3,176) = 2.97$, $R^2_p = .05$. The contrast was statistically significant for agentic applicants (difference estimate: $-.74$, 95% confidence interval: -1.18 to $-.30$), indicating that the hireability of the resolute-speaking agentic applicant in combination with a feminized job description ($M = 2.21$) was rated lower than the average hireability in the other conditions ($M = 2.95$). There was no such effect for androgynous applicants (difference estimate: $-.03$). Thus, we showed an effect of manner of speaking that was parallel to the effect that Rudman and Glick (1999, 2001) showed for applicant sex: Agentic applicants who are rated low in social skills are rated low in hireability if a job requires social skills—in their case, these were the female applicants, in our case, the resolute-speaking ones.

Gender stereotypes. Gender stereotypes were shown in implicit stereotyping (IAT internal consistency: $\alpha = .82$). As in Experiment 1, participants reacted considerably faster in the stereotype-congruent than the stereotype-incongruent task, $F(1,190) = 141.97$, $R^2_p = .43$. In an additional supplementary analysis we found an interesting priming effect of the presented applicant on implicit stereotyping. In the same ANOVA as for applicant ratings, there was an interaction of manner of speaking and applicant sex, $F(1,176) = 4.24$, $R^2_p = .02$. The only statistically significant simple main effect was the effect of manner of speaking for female

applicants, $F(1,176) = 4.84$, $R^2_p = .03$. There was more implicit stereotyping when the applicant had been a pleasant-speaking female ($M = 86$ ms) than a resolute-speaking female ($M = 56$ ms), whereas there was no such effect for male applicants ($M = 76$ ms and $M = 64$ ms). This time, controlling for experimental condition, there was no correlation between implicit stereotyping and agentic-applicant ratings ($r = .09$).

Gender stereotypes were also shown in the explicit ratings of the same characteristics as were rated for single applicants' task competence (with an internal consistency of $\alpha = .66$) and social skills ($\alpha = .79$). Task competence was rated as more characteristic of men ($M = -.48$, $SD = .49$), whereas social skills were attributed more to women ($M = .59$, $SD = .51$). A 2-level (recoded task competence stereotype vs. social skills stereotype) repeated-measures ANOVA showed a large stereotyping effect, $F(1,191) = 316.77$, $R^2_p = .62$. As an overall index of explicit stereotyping showed ($\alpha = .79$), the more explicit stereotyping participants admitted, the lower they rated the social skills of the agentic applicants ($r = -.28$, controlling for manner of speaking).

Ratings of single statements. There were no effects of experimental conditions of Phase 1 on the results of the final phase. Ratings were analyzed in 3 (type of statement: dominant vs. self-promoting vs. modest) \times 2 (applicant sex) repeated-measures ANOVAs. There was a remarkable absence of applicant sex effects except for disliking applicants with dominant statements. Specifically, the analysis on liking yielded an interaction of text content and applicant sex, $F(2,190) = 6.54$, $R^2_p = .06$. Dominant males were liked *less* than dominant females, $M = 2.45$ and $M = 2.84$ [simple main effect: $F(1,191) = 11.42$, $R^2_p = .06$], whereas there was no such simple main effect for self-promoting ($M = 3.87$ and $M = 3.81$, $F < 1$) or modest applicants ($M = 4.81$ and $M = 4.64$, $F < 2.21$). The only other effect (other $F < 1$) was a large effect of text content, $F(2,190) = 245.68$, $R^2_p = .72$, with dominant applicants liked the least. Imitation ratings showed only a main effect of text content, $F(2,190) = 122.90$, $R^2_p = .56$ (both other F s < 2.17). Applicants with dominant statements would be imitated less than self-promoting applicants ($M = 2.13$ and $M = 3.59$), and self-promoting applicants were considered more imitable than modest applicants ($M = 2.19$).

Summary. The statements of agentic applicants in Experiment 2 were crafted to be more dominant than those in Experiment 1. Still, we did not replicate a backlash effect against agentic female applicants. Instead, agentic female and male applicants were rated less hireable when the job description was feminized, due to their assumed low social skills.

In addition, as opposed to applicants who showed their task competence and were additionally dominant (agentic applicants), those who showed the same task competence, but were also communal (androgynous applicants) received higher social-skills and hireability ratings. That difference in hireability was increased with a feminized job description. An alternative means of demonstrating social skills was not in *what* was said, but *how* it was said: Applicants who spoke in a pleasant manner were, even if rated less competent than those speaking in a resolute manner, rated as more socially skilled. In addition, with a feminized job description the hireability of the resolute-speaking agentic applicants was rated lower than the hireability in all other conditions. Thus, even though there was no manipulation check on manner of speaking, the results show that the manner of speaking is a means of revealing one's social skills. In contrast to these strong effects of applicants' "gender-related behaviour", there was only one main effect of applicant sex. Female applicants appeared more task competent than male applicants, in contrast to traditionally found gender biases.

Explicit gender stereotypes predicted social-skills ratings of agentic applicants: The more gender stereotypes participants demonstrated, the lower they rated the social skills of (male or female) agentic applicants. This relationship is along the same lines as the one found with regard to implicit gender stereotyping and social-skills ratings in Experiment 1. This time, however, this relationship was not replicated on the implicit level. Possibly, given more agentic and dominant applicants (Experiment 2) blatant explicit stereotyping is related to applicant ratings; given less extreme applicants (Experiment 1) subtle implicit stereotyping is related to those ratings instead.

Final ratings of written statements showed that being dominant made both applicants appear dislikable, and males even more than females. Thus, to our participants, dominance appears to be somewhat more forgivable for females than males. At the same time, participants themselves would not make such statements.

Experiment 3

Experiments 3 and 4 were carried out in order to hunt the backlash effect by extending the boundaries of our research. Specifically, in both experiments we used instructions that were less direct than in the previous experiments in order to veil the purpose of the experiments better—if participants suspect that the research is about discrimination against female applicants, they might try to correct their biases. Therefore in Experiment 3, first we carefully probed participants about what they thought the purpose of the experiment was after the applicant ratings. Second, we manipulated the type of context in which a position was to be filled (gender neutral as in the previous experiments vs. typically male dominated). In Rudman and Glick's (1999, 2001) studies, the job (computer lab manager) was more sex-typed than the jobs in our above studies. Using sex-typed professions might make the unveiling of discriminatory processes more probable. Third, for the purpose of excluding a shifting-standards account of our findings (Biernat & Kobrynowicz, 1997), we used objective scales in addition to the subjective scales of the previous experiments. According to Biernat and colleagues, ratings on subjective scales may be subject to a shifting standard, so that a high rating may imply "for a woman, very competent" ("... but not in comparison to a man"). If such a subjective scale is replaced with an objective one, a difference between a female and male applicant could appear. For instance, a woman's height may be rated maximal (7 on a subjective scale) if she is very tall for a woman, but an objective scale that uses centimeters would reveal that she is 10 centimeters shorter than a man rated 7. Analogously, given a question like "How hireable is she?", an objective scale with the anchors "among the 10% least hireable applicants" and "among the 10% most hireable applicants" might reveal a difference in hireability that the subjective scales cannot detect. Only agentic applicants (task-competent and dominant) were presented in Experiment 3, operationalized via three female and three male voices.

Method

Participants

Participants were 144 students of various majors at a large university in former Western Germany (67% females). They ranged in age from 19 to 43 years ($M = 23$, $SD = 3$). Psychology students received course credit for participating. None had participated in Experiments 1 or 2.

Materials

The subjective scales on which applicants were rated and the IAT were identical to those used in the previous experiments. Additionally, task competence, social skills, and hireability were assessed on objective scales modelled after Biernat (2003), each with three objective questions. Specifically, participants were asked questions like “How many of ten points would the applicant get in a task competence test?”, “How many of ten social problems (e.g. team conflicts) would the applicant overcome?”, or “How many of ten interviewers would elect the applicant?”.

Job descriptions of two different economic sectors were used, both tailored to a leadership position in middle management; one signified a gender neutral context and was identical to that used in Experiment 2, the other described a position in a technical, typically male dominated context: an automobile supplier. For each economic sector a traditional and a feminized version was created. In sum, there were four different job descriptions.

The applicants replied to the same three questions as in Experiment 2. Male and female voices were rated as comparable in dominance and friendliness in a pre-study. In order to check whether the chosen job context of an automobile supplier is typically male associated, participants judged the proportion of men and women who would be chosen in the given job context (on a 5-point scale) in reality, after the manipulation check. As expected, the proportion of men who would be chosen for the job was estimated to be higher in the male dominated than the neutral job context, $t(134) = -3.82$. Subsequently, we assessed participants' intuitions about the experiment. In detail, we asked what they thought the purpose of the experiment was, which characteristics of the applicant they had concentrated on, and why they had done so. The gender stereotype IAT was administered after that, but these findings will not be reported further because replicating Experiment 2, there was no correlation between implicit stereotypes and applicant ratings (controlling for experimental condition, $|r| < .08$).³

³ Also as in Experiment 2, there was a priming effect. An ANOVA with the factors job context, job description, and applicant sex on the IAT effect showed a priming effect of applicant sex, $F(1,136) = 4.89$, $R^2_p = .04$. Interestingly, after the presentation of a (counter-stereotypical) female agentic applicant, the IAT effect was smaller than after a (stereotypical) male agentic applicant.

Procedure

With the following exception, the procedure was identical to the previous experiments. The participants were not instructed to judge applicants directly, but were instructed to rate students who had taken part in an application training, and were told that their ratings would provide very useful feedback for them (cf. Rudman & Glick, 1999). Again, after reading a job description, participants listened to segments of a job interview and were asked to form an impression of the applicant. Text content was identical to Experiment 2. All applicants displayed agentic (task-competent and dominant) behaviour. Subsequently, they rated the applicant's traits on subjective and objective scales, followed by the manipulation checks and probing for suspicion.

Results

No participant suspected that the experiment was related to applicant sex. Preliminary analyses showed there was only one effect involving voice, specifically, an interaction of job context, job description, and voice on perceived task competence, $F(2,120) = 3.39$, $R^2_p = .05$, that did not involve applicant sex and was irrelevant to our hypotheses.

All data are shown in Table 2. In order to test whether men and women were evaluated on different standards, we compared subjective and objective ratings on counter-stereotypic traits. Specifically, subjective and objective task-competence ratings were compared for female applicants, as well as subjective and objective ratings of social skills for male applicants. Both $t(71) < 1$, thus, there was no evidence for shifting standards. The statistical results of the subjective scales are reported in detail because of their better internal consistency (all α s $> .74$). Additionally, those of the objective scales where two items had to be omitted are also reported only if they diverge (final consistencies were: social skills, $\alpha = .70$, task competence, $\alpha = .77$, and hireability, $\alpha = .62$). We computed 2 (applicant sex) $\times 2$ (job description) $\times 2$ (job context) ANOVAs. As Figure 1 shows, social skills and hireability of the female applicant appear to be rated higher than those of the male, not lower, even including a male dominated job context.

Analyses of Backlash Effect in Applicant Ratings. Again, there were no interactions of applicant sex and job description on ratings of task competence, social skills, or hireability (all F s < 1). Regarding task competence, there were no statistically significant effects (all F s ≤ 2.81). As in Experiment 1, the Helmert contrast comparing subjective hireability of the female applicant, in combination with a feminized job description, to the other three conditions was not statistically significant (contrast estimate: .11, 95% confidence interval: -.21 to .43). However, the male applicant was rated as less hireable when the job description was feminized than applicants with a traditional job description (contrast estimate: -.50, 95% confidence interval: -.84 to -.17; $F[3,136] = 3.49$, $R^2_p = .07$). In addition, again, applicants were rated subjectively less hireable when the job description was feminized as opposed to traditional, $F(1,136) = 4.02$, $R^2_p = .03$, showing that agentic applicants in general, not only agentic females, appear less hireable if the job description is feminized; this effect was not significant with the less reliable objective ratings ($p = .06$).

Table 2: Mean Ratings (with standard deviations) of Applicants, Separated by Experimental Conditions, in Experiments 3 and 4

	Experiment 3: Agentic applicants only							
	Gender-neutral job context				Male-dominated job context			
	Traditional job description		Feminized job description		Traditional job description		Feminized job description	
	Male	Female	Male	Female	Male	Female	Male	Female
Subjective ratings								
Task competence	4.41 (.38)	4.43 (.39)	4.40 (.40)	4.34 (.31)	4.31 (.38)	4.32 (.37)	4.20 (.45)	4.54 (.20)
Social skills	1.91 (.36)	1.96 (.44)	1.87 (.47)	1.96 (.47)	1.95 (.38)	2.20 (.49)	1.73 (.41)	2.20 (.60)
Hirability	3.04 (.76)	3.06 (.89)	2.59 (.80)	2.76 (.91)	2.93 (.88)	3.35 (.93)	2.59 (.85)	3.31 (.58)
Objective ratings								
Task competence	7.02 (1.72)	7.39 (1.59)	7.43 (1.55)	7.24 (1.28)	7.46 (1.02)	7.74 (1.15)	7.20 (1.51)	7.54 (1.58)
Social skills	2.67 (.94)	2.89 (1.68)	2.50 (1.67)	2.94 (1.47)	2.50 (.95)	3.28 (1.68)	2.75 (1.31)	3.00 (1.33)
Hirability	5.94 (1.67)	5.86 (1.85)	5.42 (1.73)	5.17 (1.64)	6.00 (1.74)	6.81 (1.60)	5.64 (1.71)	6.25 (1.61)
Experiment 4: Agentic applicants only; male-dominated job context; subjective ratings								
	Full information				Limited information			
	Traditional job description		Feminized job description		Traditional job description		Feminized job description	
	Male	Female	Male	Female	Male	Female	Male	Female
Task competence								
No children	4.08 (.42)	4.48 (.64)	4.14 (.48)	4.22 (.57)	3.43 (.59)	4.06 (.43)	3.86 (.75)	3.90 (.46)
Children	4.42 (.33)	4.48 (.28)	3.05 (2.09)	4.30 (.22)	3.98 (.64)	3.60 (.45)	3.60 (.36)	3.38 (.30)
Social skills								
No children	2.12 (.29)	1.84 (.56)	1.84 (.53)	2.08 (.41)	3.38 (.32)	3.08 (1.00)	3.12 (.34)	3.50 (.69)
Children	2.14 (.66)	2.32 (.75)	2.47 (.60)	2.00 (.54)	3.38 (.88)	3.98 (.57)	3.90 (.61)	3.90 (.32)
Hirability								
No children	2.72 (.57)	3.40 (.72)	2.13 (.38)	2.53 (.69)	3.28 (.49)	3.93 (.64)	3.40 (.72)	3.44 (.51)
Children	3.33 (1.39)	3.47 (1.50)	2.67 (1.20)	2.75 (.74)	3.27 (.80)	3.27 (1.14)	3.58 (.74)	2.75 (.96)

Effects Involving Job Context. As to social skills, the female applicant was rated higher than the male one, $F(1,136) = 8.04$, $R^2_p = .06$. The interaction of applicant sex and job context came close to the preset criterion of statistical significance, $F(1,136) = 3.57$, $p = .06$, $R^2_p = .03$. Simple main effects analyses showed that in the male-dominated job context, the social skills of the female applicant were rated higher than those of the male, $F(1,136) = 11.17$, $R^2_p = .08$. Female applicants were also rated higher in social skills in a male-dominated as opposed to a neutral job context, $F(1,136) = 5.00$, $R^2_p = .04$.

The subjective ratings revealed that the female applicant was rated more hireable than the male, $F(1,136) = 5.79$, $R^2_p = .04$, an effect that was not present with objective hireability ratings ($F < 1$). A non-significant interaction of job context and applicant sex on subjective hireability, $F(1,136) = 3.02$, $p = .08$, was explored for theoretical reasons. In the male-dominated job context, the female applicant was rated more hireable than the male one, $M = 3.33$ vs. $M = 2.76$, simple main effect $F(1,136) = 8.79$, $R^2_p = .05$, and more hireable than the female applicant in the neutral job context, $M = 2.82$, simple main effect $F(1,136) = 4.73$, $R^2_p = .03$. An additional effect that emerged in objective ratings was that in a male-dominated context, applicants were rated more hireable than in a neutral one, $M = 6.16$ vs. $M = 5.60$, $F(1,136) = 4.17$, $R^2_p = .03$. The male-dominated context thus reinforced differences in ratings of social skills and hireability, but unexpectedly, the female applicant was favoured over the male. There were no other statistically significant effects.

Summary. Though adding a male job context and objective scales, we still did not find a backlash effect against agentic females, but instead, again evidence that agentic applicants, male and female, are considered less hireable if social skills are included in the job description. Furthermore, by adding the male dominated job context we found a pro-female gender bias for agentic applicants' social skills and hireability on subjective scales.

Experiment 3 excluded several explanations for the null effect of applicant sex in the previous experiments. First, no shifting standards effect was found. A reason for this is that the information about our applicants was not ambiguous. Biernat and Vescio (2002) reported shifting standards particularly when ambiguous information was presented. For our purposes, the crucial conclusion is that shifting standards do not seem to be a viable explanation for the lack of an applicant sex effect in our studies. Second, with more subtle instructions, careful probing which assured that participants did not guess the focus of our studies, and the inclusion of a male dominated job context, we still did not find that male agentic applicants are rated more hireable than females when the job description is feminized.

Experiment 4

We tested two further conditions that might lead to applicant sex differences in judgments of hireability: First, we tested whether a woman would appear less hireable than a comparable man as soon as she was known to have children. General gender stereotypes provide little information because they each apply to half of humankind (cf. Vonk & Olde-Monnikhof, 1998), thus conveying broad general knowledge at the cost of specific individual contents. Therefore, it is very plausible that people rely on sub-stereotypes when forming impressions. These are well-established and behaviourally relevant (see Eckes, 1997, for extended research on sub-stereotypes in Germany). Presumably, the sub-stereotype of the career woman was activated in Experiments 1-3. This sub-stereotype is very different from

the “typical woman”, and rather similar to the “typical man” as well as to the “typical manager” (Eckes, 1994). In fact, there is empirical evidence that very similar impressions are formed of career women and managers (Eckes, 2001). Additionally, in a two-dimensional space formed by *competence* and *warmth* (the core dimensions of the stereotype content model, see Fiske, Cuddy, Glick, & Xu, 2002), career women were situated among the female subtypes equivalent to where career men were situated among the male subtypes (Eckes, 2002): Both were ascribed almost maximal competence and minimal warmth. The manager was rated almost identical to them, and the “typical man” was rated rather similar, too, but not the “typical woman”. In fact, when people were asked to sort female and male subtypes into several categories at the same time, the similarities among these types could be corroborated (Carpenter & Trentham, 2001). The main result of the latter study was that sex is not a central dimension for grouping subtypes, provoking the authors to ask the question whether one should “take ‘gender’ out of gender subtypes” (p. 455).

In the present experiment, we tested whether the information that an applicant has children influences ratings of female but not male applicants (cf. Fuegen, Biernat, & Haines, 2004). Being a mother is a female sub-stereotype that is close to the typical woman. We thus assumed that the prescriptive stereotype to be warm (the one violated by agentic females) is more prescriptive for mothers than for career women (cf. Brescoll & Uhlmann, 2005, for evidence of negative attitudes towards working mothers). We therefore tested whether the backlash effect could possibly be found for agentic females who have children.

A second change in Experiment 4 was that we manipulated the amount of information given about an applicant in order to find an effect of gender stereotypes on hireability judgments. The influence of gender stereotypes on impression formation should be larger when little individuating information is present (Brewer, 1988; Fiske, Lin, & Neuberg, 1999; Fiske & Neuberg, 1990). In line with this reasoning, previous research has shown that other information than a person’s sex can be more diagnostic for gendered trait ascriptions (cf. Eagly & Steffen, 1984; Maurer & Taylor, 1994). Locksley and colleagues (Locksley, Borgida, Brekke, & Hepburn, 1980) showed that “a minimal amount of subjectively diagnostic target case information” was “sufficient to eradicate effects of stereotypes on judgments about individuals” (p. 821). According to Deaux and Lewis’ findings (1984), role information, trait information, and physical appearance information all outweighed sex in stereotyping. Pratto and Bargh (1991) demonstrated that behavioural information clearly related to gender leads to gender-stereotypic impressions, regardless of the sex of the person rated (also see Gerber, 1988). In sum, it has been shown that males are ascribed typically masculine traits, and females, typically feminine ones, if nothing but a person’s sex is known. As soon as additional gender-role related information is presented, inferences consistent with that—more diagnostic—information are drawn. Thus, a female applicant might appear less hireable than a comparable male if there was only very limited information available about her. In the limited information condition of Experiment 4, the applicant’s self-presentation stopped after they had spoken the first few words, so that nothing is known about text content. If more individuating information is needed from a female as opposed to a male applicant for forming the impression that she is hireable, she should be judged lower than him in the limited information condition.

The experiment was identical to the agentic, pleasant-speaking applicant conditions of the previous experiments in all other respects.

Method

Participants

Participants were 76 students of various majors (excluding psychology) at a large university in former Eastern Germany (83% females). They ranged in age from 18 to 39 years ($M = 21$, $SD = 3$) and received course credit or candy for participating.

Materials and procedure

The materials and procedure were identical to those in the previous experiments. In particular, the alleged study purpose, the subjective scales on which participants rated applicants' task competence, social skills, and hireability, as well as the traditional and feminized job descriptions were identical to those used in Experiment 3.

First, the instructions and one of the job descriptions (traditional or feminized) of the male dominated job context of Experiment 3 were presented. In order to provide a suitable context for the parental status manipulation, participants then received a few pieces of information about the fictitious applicants: "The applicant is Marion (Mario) Kraft, 28 years old, and has just received her (his) major in business studies from the University of Leipzig. Her (His) hobby is sports, and she (he) has no (two) children." After that they listened to the agentic (task-competent and dominant) statements of the pleasant-speaking male or female voice from Experiment 2. The amount of available information was manipulated by presenting either two detailed answers in an alleged job interview (henceforth: full information) or interrupting the interview after the first few words (henceforth: limited information). Before applicant ratings, in the limited information condition participants were explained that they were in a "first impression" condition. After the applicant ratings and the manipulation check with regard to the job description, as an additional manipulation check, participants were asked to judge whether the applicant had two children, among four distractor questions. Only 3 answered incorrectly on the parental status, and excluding them did not change the pattern of findings below. Measures of gender stereotypes were omitted.

Results

For ratings on task competence, social skills, and hireability (all $\alpha s > .79$), we computed separate 2 (applicant sex) \times 2 (job description) \times 2 (amount of information received: full vs. limited information) \times 2 (parental status: no children vs. two children) ANOVAs.

Analyses of backlash effect in applicant ratings. As Figure 1 shows, female and male agentic applicants were rated comparably low in hireability when the job description was feminized. Again, there was no interaction of job description and applicant sex on ratings of task competence, social skills, or hireability (all $F s < 1.16$). Also, Helmert contrasts testing hireability judgments of the female applicant when the job description was feminized against all other applicants yielded no statistically significant effects ($F < 2$). Additionally, applicants were generally rated more hireable when the job was described traditionally as opposed to feminized, $M = 3.32$ and $M = 2.88$, respectively, $F(1,59) = 4.27$, $R^2_p = .07$, again

replicating the general effect that female and male agentic applicants appear less hireable if social skills are part of the job description.

Effects of full versus limited information. Applicants' task competence was rated higher after full rather than limited information had been provided, $M = 4.17$ and $M = 3.72$, respectively, $F(1,60) = 7.27$, $R^2_p = .11$. This shows that the materials we prepared indeed conveyed task competence. There were no other statistically significant effects for task competence (all $F_s < 3.25$).

As we had expected, social skills of applicants were rated higher if participants had only limited information than if they received full information, $M = 3.51$ and $M = 2.08$, respectively, $F(1,59) = 104.53$, $R^2_p = .64$, showing that agentic applicants actually conveyed low social skills. Finally, hireability ratings were higher when limited information about an applicant was given than full information, $M = 3.37$ and $M = 2.89$, respectively, $F(1,59) = 5.65$, $R^2_p = .09$. Apparently, targets behaved too dominant to appear very hireable in spite of their task competence (for hireability, all other $F_s < 2.52$, including all effects of applicant sex).

Effects of parental status. Interestingly, applicants were rated higher in social skills if they had children than if they had none, $M = 3.02$ and $M = 2.58$, respectively, $F(1,59) = 7.81$, $R^2_p = .12$. There were no other effects on social skills judgments (all $F_s < 1$) except for an unexpected interaction of applicant sex, job description, and parental status on social skills, $F(1,59) = 4.78$, $R^2_p = .08$, irrelevant to our hypotheses.⁴

Summary. We found no hiring bias against female applicants when the job description was feminized. Even if only very limited information was presented about a female or male applicant, and even if the fictitious applicants have children, the hireability of the female was not rated lower than that of the male. Our manipulations did show, however, that the responses of our applicants indeed appeared very task competent but inappropriate: In the limited information condition applicants appeared less task competent, but more socially skilled and more hireable. Again, as in the previous experiments, we found that both male and female dominant applicants were considered less hireable if the job description was feminized.

General discussion

Rudman and Glick (1999, 2001) reasoned that female applicants who convey typically male agency violate the prescriptive stereotype that females should be nice. This violation should be used against these female applicants only when social skills are part of the job description. This interaction effect was shown using videos of alleged job applicants: The social skills of the female applicants were rated lower than those of the male applicants, and when the job description was feminized, the agentic females were rated less hireable than the agentic males.

⁴ Tests of simple main effects revealed that given a *female applicant* and a *traditional* job description, applicants who had children were rated more socially skilled than those who had no children, $F(1,28) = 5.74$, $R^2_p = .17$. A similar pattern was found given *male applicants* and a *feminized* job description: Those who had children were rated more socially skilled than those who had none, $F(1,31) = 6.78$, $R^2_p = .18$; both other simple main effects were not significant (both $F_s < 1$).

In the present set of experiments, in which female and male voices were used, agentic females' social skills were not rated lower than agentic males'. In addition, in contrast to traditional gender stereotypes, the task competence of the females was not rated lower than that of the males. Consequently, we observed no hiring bias against female applicants, neither with a traditional nor with a feminized job description; neither when they were mothers nor when they were not. At the same time, we demonstrated expected effects of text content: Agentic applicants were rated higher in task competence, but lower in social skills than communal or androgynous applicants. A similar pattern was found for the amount of information given: When the whole agentic and dominant text had been presented, applicants were rated higher in task competence and lower in social skills than when hardly any information had been given. Also, more social skills were attributed to parents compared to people without children. Finally, we found expected effects of manner of speaking: Applicants speaking in a resolute way were rated higher in task competence and lower in social skills than those speaking in a friendly manner.

In order to analyze our data with the highest available statistical power, we did a combined analysis of all ratings of agentic applicants ($N = 457$) that also included participant sex as a factor; 2 (participant sex) \times 2 (job description) \times 2 (applicant sex) \times 4 (experiment) ANOVAs yielded neither interactions of applicant sex and job description nor interactions of participant sex and applicant sex on task competence, social skills, or hireability (all F s < 1). Thus, there was no backlash effect against female agentic applicants.

Why did we not find a backlash effect?

As the effect of manner of speaking shows, studies with only a few applicants of each sex are difficult to interpret because some differences between the females and males may be present in the way they presented themselves (cf. Wells & Windschitl, 1999). Thus, one possible explanation for Rudman and Glick's (1999, 2001) findings is that by chance, their female applicants appeared less socially skilled than the male ones, and any applicant who appears low in social skills is rated less hireable when the job description is feminized as opposed to traditional. In other words, disliked applicants are discriminated against if one finds a reason for justifying that discrimination (cf. Uhlmann & Cohen, 2005). In line with this reasoning, the hireability of agentic applicants in our experiments was rated lower when the job description was feminized rather than traditional, and Experiment 2 showed that these applicants were indeed disliked. In addition, the social skills of resolute-speaking applicants were rated lower than those of pleasant-speaking applicants, and this resulted in lower hireability ratings when the job description was feminized.

Note that except for Experiment 4, we did not assign first names to our applicants. This knocks out a confound inherent in the selected first names (Kasof, 1993) as an alternative explanation for our findings. What we cannot rule out is the idea that a backlash effect only occurs in the presence of visual signals showing what an agentic woman looks like.

Other likely explanations for the difference in findings between the U.S.-based previous studies and the present research conducted in Germany could be changes in roles and norms in the past decade or cultural differences. In fact, an explanation along these lines is corroborated by a recent study in which the backlash effect was not replicated with female participants in the Netherlands, even though Rudman and Glick's original videotapes were used

(Rudman, 2008, personal communication). Hostility towards agentic women could have waned across the past decade because changing female role norms now incorporate agentic behaviour (Diekmann & Goodfriend, 2006). With regard to cultural differences, gender stereotypes and gender empowerment indices are similar in these three countries, so the question remains what the cultural difference could be. Our speculation is: The degree of self-promotion and dominance that seems tolerable for the American male applicant appeared quite inappropriate to the German participants. In other words, a cultural difference might not be rooted in the degree to which a female can behave dominant and still appear socially skilled, but in the acceptable behaviour for a European male: It is impossible to behave as dominant as an American male and still appear socially skilled.

Some of our data corroborate that explanation. First, across all our experiments we replicated that both female and male agentic applicants were rated lower in hireability if social skills were added to the job description (replicating Steffens & Mehl, 2003). Moreover, our agentic applicants appeared less hireable than androgynous applicants (Experiment 2) and also less hireable than applicants about whom hardly anything was known (Experiment 4), despite their high competence. In other words, they conveyed something that markedly reduced their hireability, and this appears to be their dominant behaviour. The finding of Experiment 2 that our participants would not make such dominant statements in a job interview themselves is in line with this explanation.

What we found most striking across the current set of experiments was how few applicant sex effects we found. Of course, this is not to say that applicant sex does not influence hireability decisions. But our participants, samples of students—most of them female—appear to rely very little on gender stereotypes when forming impressions of applicants. Gender biases may have become more subtle. For instance, Uhlmann and Cohen (2005) reported that participants redefined the criteria for success at a job that was extremely stereotypical of their own sex as requiring the specific credentials of the ingroup candidate. A finding in their data on which Uhlmann and Cohen put little stress is in line with our findings: Male and female applicants' features were rated similarly (for instance, an "educated female" was rated as educated as an "educated male" was). Only when asked how important being educated was for the job, did discrimination occur.

Gender stereotypes & competence ascriptions

Except for a reversed gender bias in one experiment, we found that task competence was similarly ascribed to the male and female candidates. Should task competence ascriptions depend on applicant sex? In the 1970ies, a gender bias was found using the Goldberg paradigm: The quality of identical texts was rated higher when they were attributed to a male rather than female author (Goldberg, 1968). Later studies used biographies, job applications, resumes, and behaviour as stimulus materials in that paradigm. The gender bias was not always replicated and its overall size was small, $d = .07$ (for a review, see Swim et al., 1989). Accordingly, in the context of personnel selection some meta analyses concluded that there are applicant sex effects, albeit small ones (Davison & Burke, 2000; Eagly et al., 1992; Olian, Schwab, & Haberfeld, 1988), whereas other authors stressed that the evidence for sex discrimination in this context is meager (Campion & Arvey, 1989; Graves, 1999; Powell, 1987).

The effect sizes reported by Swim et al. (1989) proved to be heterogeneous, a finding that points at moderators of the effect (see Banaji & Greenwald, 1994, for a discussion). For instance, the effect size was larger if only the name of the target person was provided than if more information was given. In line with this, Heilman (1984) reported that differential treatment of male and female applicants was diminished if highly job-relevant information about them was presented. Similarly, fictitious male employees were rated significantly more competent than female employees when information on former success was ambiguous, but equally when the employees had a record of clear previous success (Heilman, Wallen, Fuchs, & Tamkins, 2004). Taken together, task competence should be attributed to a male as well as to a female applicant when enough information is presented about them.

An international research program pursued by Diekmann and colleagues (Diekmann & Eagly, 2000; Wilde & Diekmann, 2005) also indicates that task competence, as one aspect of agency, may be equally attributed to female and male applicants. According to their findings, along with the increasing participation of females in all kinds of social roles (see Eagly & Steffen, 1984, for more theoretical background), traditionally masculine attributes are increasingly ascribed to females as well. This is in line with newer studies that show no general automatic gender stereotypes with regard to the competence dimension, a finding that is assumed to follow from women's increasing participation in competence-associated roles (Ebert, Steffens, Goergens, & Kroth, 2008). Related evidence shows that exposure to successful women leaders diminishes stereotyping and discrimination (Dasgupta & Asgari, 2004; Heilman & Martell, 1986; also cf. Pazy, 1986). Conceivably, in so far as gender stereotypes have eroded with regard to task competence or agency, females and males should not be evaluated differently anymore.

Given these findings, one should expect that it is not females in general who are attributed fewer leadership qualities as compared to males, but instead, typical females or, even more generally, people with typically feminine attributes; for instance, female or male applicants who had applied a typically female perfume (Sczesny & Stahlberg, 2002), or those with typically female physical features (Heilman & Stopeck, 1985; Riehle, 1996; cf. also Sczesny & Kühnen, 2004). In line with these findings on gender stereotyping in general, target sex should play only a minor, if any, role in the assessment of leaders' traits. Indeed, Vecchio (2003) asked researchers to directly compare the influences of biological sex and gender-typical behaviour on leaders' assessment. Our findings confirm that the effect of target sex on applicant ratings of competence, social skills, and hireability is much smaller than that of other applicant features.

At the same time, a general gender stereotype—a male-agency, female-social skills association—still seems to exist according to our data. If nothing is known about a person but their sex, men are believed to be relatively agentic, women communal. This was evident in implicit and explicit measures of gender stereotypes. Strong implicit gender stereotyping was shown in Experiments 1-3. In addition, when we asked participants in Experiment 2, they were also willing to admit these stereotypes on explicit ratings, when they ascribed task competence more to males, and social skills, more to females. However, individual agentic females and males were perceived quite similarly, as were communal or androgynous males and females.

Recently, Gill (2004) reported that prescriptive stereotyping (“women *should be* gentle”) can lead to gender bias even if behavioural information deters descriptive stereotyping. According to his analyses, individual differences among male raters in prescriptive stereotypes

predicted bias against women. Because there is not even a tendency in our experiments for higher task-competence ratings of male as compared to female applicants, such a bias is unlikely in our study.

Gender stereotypes & applicant ratings

An interesting side note: Gender stereotypes predicted the social-skills ratings of male and female agentic applicants, that is, applicants for whom no information regarding social skills was given. We showed this for explicit stereotypes in Experiment 2 and for implicit stereotypes in Experiment 1. This finding is, first, an indicator of IATs' power to predict behaviour. Second, we consider the relationship that was found regardless of applicant sex as another indicator of the minor influence of target sex during impression formation. "Career women" are perceived as similar to "career men" and managers (Eckes, 1994), and people who hold strong gender stereotypes expect few social skills of males and females presenting themselves in an agentic way.

An additional side note is that we observed priming effects in Experiments 2 and 3. Implicit gender stereotyping was stronger after stereotype-congruent rather than incongruent applicants had been presented. This is in line with previous research showing that implicit attitudes are context-sensitive (for a review, see Blair, 2002). More closely related to the present research, stereotyping of women as weak and men as strong was reduced after participants had imagined a strong woman (Blair, Ma, & Lenton, 2001). Apparently, a more subtle activation of counter-stereotypes is sufficient for reducing implicit stereotyping: Judging a woman who violates gender stereotypes by speaking in a dominant, assertive way (Experiment 2) or simply judging an agentic woman (Experiment 3) before the stereotype-assessment task mitigated implicit gender stereotypes.

Conclusion

There is a lot of evidence in the recent literature of prejudice and discrimination against women as leaders (e.g., Biernat & Kobrynowicz, 1997; Rudman, 1998). However, it has also been previously concluded that applicant sex effects are rather small in comparison to other differences between applicants (e.g., Olian et al., 1988). We found that after phone interviews, students rated both females and males, despite their competence, low in hireability when they appeared dominant and when social skills were part of the job description. It is, of course, impossible to substantiate the claim empirically that applicant sex does not affect hireability judgments. There may always be other circumstances under which effects are obtained. Such moderators may be context factors (Woehr & Roch, 1996), or, in light of stereotype activation and application (Kunda & Spencer, 2003): "if decision makers lack the time, opportunity, or inclination to carefully process information, the likelihood of gender bias remains high" (Martell, 1996, p. 162). Hiring biases might be found given the more complex interactions that precede hiring decisions in reality, and given gender-correlated behavioural differences in such interactions. Still, there was a remarkable absence of gender bias against females in the present experiments with mostly female students as participants (also see Steffens & Mehl, 2003). Therefore, we conclude that what applicants say in a job

interview, and how they say it, is fundamentally more important than whether they are male or female.

Author notes

This research was supported by a grant from the Deutsche Forschungsgemeinschaft (Ste 938/6-1). We are much indebted to Laurie Rudman, Anna Günster, Jutta Hartmann, and Christina Hoffmann for great help in all stages of this research project, to Felicia Pratto and Kay Deaux for their suggestions concerning the design of Experiment 4, and to Anika Fäsche, Juliane Hartmann, Tamara Rakic, and Amir Rosenmann for very valuable comments on earlier versions of this article, and Erin Marie Thompson for language corrections.

References

- Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs*, *35*, 492-503.
- Allport, G. W., & Cantril, H. (1934). Judging personality from voice. *Journal of Social Psychology*, *5*, 37-55.
- Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *Journal of Social Psychology*, *99*, 207-220.
- Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, III & J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder. Science conference series* (pp. 117-150). Washington, DC, USA: American Psychological Association.
- Banaji, M. R., & Greenwald, A. G. (1994). Implicit stereotyping and prejudice. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium, Vol. 7* (pp. 55-76). Hillsdale, NY, USA: Erlbaum.
- Berry, D. S. (1990). Vocal attractiveness and vocal babyishness: Effects on stranger, self, and friend impressions. *Journal of Nonverbal Behavior*, *14*, 141-153.
- Berry, D. S. (1991). Accuracy in social perception: contributions of facial and vocal information. *Journal of Personality and Social Psychology*, *61*, 298-307.
- Berry, D. S. (1992). Vocal types and stereotypes: Joint effects of vocal attractiveness and vocal maturity on person perception. *Journal of Nonverbal Behavior*, *16*, 41-54.
- Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, *72*, 544-557.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242-261.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828-841.
- Borgida, E., Hunt, C., & Kim, A. (2005). On the use of gender stereotyping research in sex discrimination litigation. *Journal of Law and Policy*, *13*, 613-628.
- Brescoll, V. L., & Uhlmann, E. L. (2005). Attitudes toward traditional and nontraditional parents. *Psychology of Women Quarterly*, *29*, 436-445.

- Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. J. Wyer (Eds.), *Advances in social cognition, Volume 1: A dual process model of impression formation* (pp. 1-36). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.
- Campion, J. E., & Arvey, R. D. (1989). Unfair discrimination in the employment interview. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 61-73). Thousand Oaks, CA, US: Sage Publications, Inc.
- Carpenter, S., & Trentham, S. (2001). Should we take "gender" out of gender subtypes? The effects of gender, evaluative valence, and context on the organization of person subtypes. *Sex Roles, 45*, 455-480.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised ed.). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.
- Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40*, 642-658.
- Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*, 225-248.
- Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology, 46*, 991-1004.
- Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin, 26*, 1171-1188.
- Diekman, A. B., & Goodfriend, W. (2006). Rolling with the changes: A role congruity perspective on gender norms. *Psychology of Women Quarterly, 30*, 369-383.
- Dubno, P. (1985). Attitudes toward women executives: A longitudinal approach. *Academy of Management Journal, 28*, 235-239.
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573-598.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111*, 3-22.
- Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46*, 735-754.
- Ebert, I. D., Steffens, M. C., Goergens, K., & Kroth, A. (2008). Competent and cold men – competent and warm women: Contemporary implicit gender stereotypes. *Manuscript in preparation*.
- Eckes, T. (1994). Features of men, features of women: Assessing stereotypic beliefs about gender subtypes. *British Journal of Social Psychology, 33*, 107-123.
- Eckes, T. (1997). *Geschlechterstereotype: Frau und Mann in sozialpsychologischer Sicht (Gender stereotypes: Woman and men from a social psychology viewpoint)*. Pfaffenweiler: Centaurus.
- Eckes, T. (2001). Ambivalenter Sexismus und die Polarisierung von Geschlechterstereotypen (Ambivalent sexism and the polarization of gender stereotypes). *Zeitschrift für Sozialpsychologie, 32*, 235-247.
- Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles, 47*, 99-114.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

- Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231-254). New York, NY, USA: The Guilford Press.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 23* (pp. 1-74). New York, NY, USA: Academic Press.
- Fuegen, K., Biernat, M., & Haines, E. (2004). Mothers and fathers in the workplace: How gender and parental status influence judgments of job-related competence. *Journal of Social Issues, 60*, 737-754.
- Gerber, G. L. (1988). Leadership roles and the gender stereotype traits. *Sex Roles, 18*, 649-668.
- Gill, M. J. (2004). When information does not deter stereotyping: Prescriptive stereotyping can foster bias under conditions that deter descriptive stereotyping. *Journal of Experimental Social Psychology, 40*, 619-632.
- Goldberg, P. A. (1968). Are women prejudiced against women? *Transaction, 5*, 28-30.
- Goldman, M., & Fordyce, J. (1983). Prosocial behavior as affected by eye contact, touch, and voice expression. *Journal of Social Psychology, 121*, 125-129.
- Graves, L. M. (1999). Gender bias in interviewers' evaluations of applicants: When and how does it occur? In G. N. Powell (Ed.), *Handbook of gender and work* (pp. 145-164). Thousand Oaks, CA, US: Sage Publications, Inc.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.
- Haslett, B. J., Geis, F. L., & Carter, M. R. (1992). *The organizational woman: Power and paradox*. Norwood, NJ: Ablex.
- Heilman, M. E. (1983). Sex bias in work settings: The Lack of Fit model. *Research in Organizational Behavior, 5*, 269-298.
- Heilman, M. E. (1984). Information as a deterrent against sex discrimination: The effects of applicant sex and information type on preliminary employment decisions. *Organizational Behavior & Human Performance, 33*, 174-186.
- Heilman, M. E., & Martell, R. F. (1986). Exposure to successful women: Antidote to sex discrimination in applicant screening decisions? *Organizational Behavior & Human Decision Processes, 37*, 376-390.
- Heilman, M. E., & Stopeck, M. H. (1985). Being attractive, advantage or disadvantage? Performance-based evaluations and recommended personnel actions as a function of appearance, sex, and job type. *Organizational Behavior & Human Decision Processes, 35*, 202-215.
- Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*, 416-427.
- Juodvalkis, J. L., Grefe, B. A., Hogue, M., & Svyantek, D. J. (2003). The effects of job stereotype, applicant gender, and communication style on ratings in screening interviews. *The International Journal of Organizational Analysis, 11*, 67-84.
- Kasof, J. (1993). Sex bias in the naming of stimulus persons. *Psychological Bulletin, 113*, 140-163.
- Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*, 522-544.

- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *39*, 821-831.
- Martell, R. F. (1996). What mediates gender bias in work behavior ratings? *Sex Roles*, *35*, 153-169.
- Maurer, T., J., & Taylor, M. A. (1994). Is sex by itself enough? An exploration of gender bias issues in performance appraisal. *Organizational Behavior & Human Decision Processes*, *60*, 231-251.
- Olian, J. D., Schwab, D. P., & Haberfeld, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. *Organizational Behavior and Human Decision Processes*, *41*, 180-195.
- Page, R. A., & Balloun, J. L. (1978). The effect of voice volume on the perception of personality. *Journal of Social Psychology*, *105*, 65-72.
- Pazy, A. (1986). The persistence of pro-male bias despite identical information regarding causes of success. *Organizational Behavior & Human Decision Processes*, *38*, 366-377.
- Powell, G. N. (1987). The effects of sex and gender on recruitment. *Academy of Management Review*, *12*, 731-743.
- Pratto, F., & Bargh, J. A. (1991). Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload. *Journal of Experimental Social Psychology*, *27*, 26-47.
- Riehle, D. (1996). *Personalbeurteilung auf dem Hintergrund sozialer Kategorisierungsprozesse (Personnel evaluation based on social categorization)*. Lengerich: Pabst.
- Rodler, C., Kirchler, E., & Hölzl, E. (2001). Gender stereotypes of leaders: An analysis of the contents of obituaries from 1974 to 1998. *Sex Roles*, *45*, 827-843.
- Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*, 629-645.
- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*, 1004-1010.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*, 743-762.
- Sczesny, S. (2003). Führungskompetenz: Selbst- und Fremdwahrnehmung weiblicher und männlicher Führungskräfte (leadership competence: Self and other perception of female and male leaders). *Zeitschrift für Sozialpsychologie*, *34*, 133-145.
- Sczesny, S., & Kühnen, U. (2004). Meta-cognition about biological sex and gender-stereotypic physical appearance: Consequences for the assessment of leadership competence. *Personality and Social Psychology Bulletin*, *30*, 13-21.
- Sczesny, S., & Stahlberg, D. (2002). The influence of gender-stereotyped perfumes on leadership attribution. *European Journal of Social Psychology*, *32*, 815-828.
- Steffens, M. C., & Mehl, B. (2003). Erscheinen „Karrierefrauen“ weniger sozial kompetent als „Karrieremänner“? – Geschlechterstereotype und Kompetenzzuschreibung (Do “career women” appear less socially competent than “career men”? Gender stereotypes and competence ascription). *Zeitschrift für Sozialpsychologie*, *34*, 173-185.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, *105*, 409-429.
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, *16*, 474-480.
- Vecchio, R. P. (2003). In search of gender advantage. *The Leadership Quarterly*, *14*, 835-850.

- Vonk, R., & Olde-Monnikhof, M. (1998). Gender subgroups: Intergroup bias within the sexes. *European Journal of Social Psychology, 28*, 37-41.
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115-1125.
- Wilde, A., & Diekmann, A. B. (2005). Cross-cultural similarities and differences in dynamic stereotypes: A comparison between Germany and the United States. *Psychology of Women Quarterly, 29*, 188-196.
- Woehr, D. J., & Roch, S. G. (1996). Context effects in performance evaluation: The impact of rater sex and performance level on performance ratings and behavioral recall. *Organizational Behavior and Human Decision Processes, 66*, 31-41.
- Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior, 13*, 67-82.