FlashReport

Building blocks of bias: Gender composition predicts male and female group members' evaluations of each other and the group

Tessa V. West [a],*, Madeline E. Heilman [a], Lindy Gullett [a,1], Corinne A. Moss-Racusin [b,1], Joe C. Magee [a]

[a] *New York University, NY, USA*
[b] *Yale University, CT, USA*

ARTICLE INFO

ABSTRACT

The present research examined how a group's gender composition influences intragroup evaluations. Group members evaluated fellow group members and the group as a whole following a shared task. As predicted, no performance differences were found as a function of gender composition, but judgments of individuals' task contributions, the group's effectiveness, and desire to work with one's group again measured at a 10-week follow-up were increasingly negative as the proportion of women in the group increased. Negative judgments were consistently directed at male and female group members as indicated by no gender of target effects, demonstrating that men, simply by working alongside women, can be detrimentally affected by negative stereotypes about women. Implications for gender diversity in the workplace are discussed.

Introduction

Research consistently indicates that gender affects judgments in work settings. Perceptions of women's incompetence and negative evaluations of their performance arise from the belief that women are deficient in the male stereotyped agentic attributes (e.g., ambitious, competitive) that are required for success in male sex-typed roles (Eagly & Karau, 2002; Heilman, 2001; Rudman, Moss-Racusin, Phelan, & Nauts, 2012). This negative effect of gender stereotypes can persist even in the presence of disconfirming behavioral evidence (Foschi, 1996). Judgments of work groups also are affected by gender such that gender-diverse task groups are perceived as less effective than are task groups with more men (Baugh & Graen, 1997). The study reported here addresses how gender composition of task groups affects *intragroup* processes; namely, members' evaluations of male and female group members and the group as a whole. Unlike past research, our focus is not on third party evaluations of group members, but rather on group members' evaluations of each other.

It would not be surprising if the performance of female group members was negatively evaluated within work groups composed of more women than men, particularly on male gender-typed tasks for which there is a lack of fit between the attributes women are thought to embody and the attributes believed to be required for success. However, we propose that a group composed of more women than men can also negatively affect how *male* group members are evaluated. Two different streams of research lead us to this prediction.

For one, when completing a male-typed task for which gender stereotypes are relevant, it is possible that in female-dominated groups, stereotypes about women's deficiencies in agentic attributes "leak" into evaluations of men when those men are working interdependently with women, thereby detrimentally affecting views of their competence and performance effectiveness. This idea is consistent with research demonstrating that when there is a high degree of entitativity between stigmatized and non-stigmatized persons (e.g., shared outcomes and common goals), a "stigma-by-association" effect occurs whereby non-stigmatized persons are ascribed stereotypical traits of the stigmatized persons (Pryor, Reeder, & Monroe, 2012). It is also consistent with the finding that men are perceived to be deficient in agentic attributes when working in female-dominated occupations and job contexts (Heilman & Wallen, 2010).

Alternatively, the proportion of stereotyped individuals in a group (in this case women) might be an emergent property of the group itself, which may color the perceptions and experiences of its members. As the number of women in a group increases, the views of the group as a whole may change in a way that is consistent with negative stereotypes about women in the workplace, and filter down to perceptions of group members—whether male or female. This idea is consistent with prior work demonstrating that group-level category memberships can shape individual group members' judgments of the group as a whole and each other, above and beyond those individuals' own category memberships (Gaertner & Dovidio, 2000; Hewstone & Brown, 1986). However, whereas that research has historically been concerned with examining the evaluative benefits of positively valued group-level category membership on

perceptions of outgroup members (Gaertner & Dovidio, 2000), the current research is focused on the potential evaluative costs of negatively valued group-level category membership on perceptions of outgroup members; in this case men in a predominantly female group.

With these ideas as a backdrop, we propose that the greater the number of female members in a group working on a male gender-typed task, the more the group members will make negative judgments about each others' performance, and these negative judgments will be directed not only at female group members but also at male group members. Our objective is to extend prior work on the effects of gender stereotypes to the workgroup level, potentially demonstrating an unrecognized consequence of gender diversity within task-focused settings—that even group members who are positively stereotyped can be detrimentally affected when affiliated with a negatively stereotyped group.

Overview

In this study, group members worked interdependently toward a common goal in mixed-gender teams and were evaluated objectively as a group on a task for which men are believed to outperform women. Consistent with past research (Myaskovsky, Unikel, & Dew, 2005), we did not expect objective performance to be influenced by the gender composition of the group. However, we did predict gender composition to affect subjective judgments such that women's *and* men's task contributions would be perceived more negatively in groups with a higher proportion of women. Furthermore, we expected individuals in groups with a higher proportion of women to be more dissatisfied with the group as a whole and to express less interest in working with the group again.

Method

Participants

Participants were 110 (71 female) graduate students enrolled in four sections of an introductory management course. They were randomly assigned to 22 5-person groups (*M* age = 26.41 years; 56% White)[2].

Procedure

Our independent variable was the number of women per group, which ranged from two to four (9% had two women; 59% had three women; 32% had four women). All groups were the same size, so we refer to this variable as *proportion female*. During the second meeting of the semester, participants were randomly assigned to groups and were told that their goal as a group was to build a replica of a complex model made of Legos™. This task requires contributions from all members (Heath & Staudenmayer, 2000), and it has been shown that participants believe the task is better suited to men's than women's abilities (Loyd, White, & Kern, submitted for publication). Groups had 30 min to plan a strategy, followed by another 30 min to build their replica. When groups believed they had completed an exact replica, they presented their model to a judge. If the model was incorrect, the judge rejected it but did not give the group any information about the defect. This procedure continued until the group assembled an exact replica of the model.

After the task, participants made round-robin evaluations of the other group members' task contributions and of the overall effectiveness of their group. To explore whether perceptions of group members' competence persisted beyond this immediate task setting, ten weeks later participants reported their interest in working with their group again on a graded group project.

Measures

Participants reported all ratings on a 1 (*disagree strongly*) to 7 (*agree strongly*) scale. Our measure of objective performance was the number of times the group's model was rejected by the judge for having a defect (range: 0–4) (Heath & Staudenmayer, 2000).

We measured each member's task contribution with round-robin ratings of the extent to which each group member *contributed to*, *was focused on*, *was competent at*, and *helped the team complete* the task ($\alpha = .89$). We also measured perceptions of group effectiveness with the following three items: "The team as a whole worked well together", "My team was good at coordinating the work of all of the members", and "I would like to work with the team in the future" ($\alpha = .89$).

Ten weeks later, we measured participants' desire to work together again with one 7-point scale item on a web-based questionnaire: "To what extent would you be willing to work with your Legoperson team on a graded group project?" In the intervening weeks, participants worked in groups on a graded team project that constituted 40% of their grade, so they were aware of the high-stakes nature of a project of this kind. We made sure that no participants worked with their Lego task group members on this team project. At this point, we also measured their recollection of how well their group performed on the Lego task, on a scale ranging from 1 (*really poorly*) to 7 (*really well*).

Results

We treated gender composition of the group (*proportion of women*) as a continuous variable (mean centered) in all models. Analyses treating proportion of women as a categorical variable—comparing groups with two women to those with three women, and comparing groups with three women with those with four women—yielded the same pattern of effects. For judgments of the group as a whole (group effectiveness and desire to work together again), we used multilevel modeling to account for non-independence of data reported by group members within the same group. We analyzed data for round-robin ratings of task contribution using Kenny and Livi's (2009) multilevel modeling strategy for the analysis of Social Relations Model data.

Dependent measures

Correlations between the dependent measures are presented in Table 1.

Performance
As expected, the proportion of women in the group did not predict performance, $t(1) = .343$, $p = .735$.[3]

Task contribution
In addition to including the proportion of women in the group as a predictor, we included gender of the target, gender of the perceiver and the interaction between target gender and perceiver gender to test whether men were perceived differently than women, whether men's judgments differed from women's, and whether people judged others who were of the same gender as themselves differently than those of the other gender. We also included the two-way interactions

[3] Proportion female also did not predict performance, conceptualized as rejections relative to other teams in the class, $p = .646$.

Table 1
Correlations between dependent measures. Group-level means were created for task contribution, group effectiveness, and desire to work together again.

	Performance	Task contribution	Group effectiveness	Desire to work together again
Performance		−.550**	−.299	−.188
Task contribution			.814***	.780***
Group effectiveness				.717***

** $p < .01$.
*** $p < .001$.

Fig. 1. Adjusted means for ratings of individuals' task contribution.

of proportion of women by gender of the target to test whether men and women were rated differently in groups with different proportions of women, and the proportion of women by gender of the perceiver to test whether there were differences in men's versus women's ratings in groups that had different proportions of women.

Results indicate that as the proportion of women increased, group members' judgments of each others' task contributions decreased, $t(131.95) = -2.23$, $p = .027$ (see Fig. 1, which displays adjusted means). Gender of the target was not significant, nor was gender of the perceiver, or their interaction ($ps > .75$). The 2-way interactions of proportion of women in the group with perceiver gender and with target gender were also not significant, $ps > .88$. Thus, in groups with a greater proportion of women, all team members rated each other as contributing less to the task, regardless of the gender of the person being judged or the gender of the person making judgments.

Group effectiveness

Consistent with findings for task contributions, as the proportion of women in the group increased, judgments of group effectiveness decreased, $t(18.79) = -2.27$, $p = .039$ (see Fig. 2). We also included the perceiver gender main effect and its interaction with the proportion of women, neither of which were significant, $ps > .47$.

Desire to work together again

Sixty-five percent of participants responded to this item asked ten weeks after the experimental task. Data were missing at random with respect to variables of interest (proportion of women in the group and perceiver gender, $ps > .516$). Because a significant amount of time elapsed between the task and the follow-up questionnaire, recall of task performance was adjusted for by including members' recollections of their group's performance as a predictor. As with analyses for group effectiveness, the interaction between gender of the perceiver and proportion of women was also included.

Participants who remembered their group performing better were more willing to work with their groups again on a group project, $t(36.93) = 3.84$, $p = .001$. Adjusting for this effect, groups with higher proportion of women were less interested in working together again, $t(17.86) = -2.62$, $p = .018$ (see Fig. 3). There was no main effect of gender of the perceiver or proportion of women by perceiver gender interaction, $ps > .075$.

Discussion

Although there was no evidence that actual performance was influenced by gender composition of the group, the study's results indicate that group members' task contributions were evaluated more negatively by other group members as the proportion of women in the group increased. This occurred regardless of whether the group members being evaluated were male or female, suggesting that gender stereotypes about perceived deficiencies in women's performance-related agentic attributes are applied to men as well as women when they work in primarily female contexts. Thus, the present work shows that men can be evaluated just as negatively as women, simply by
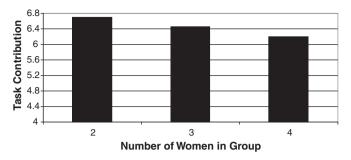
their association with women. We also found that individuals perceived their group as a whole more negatively, and were less interested in future collaborations, as the proportion of women in the group increased. These results are consistent with previous research demonstrating negativity toward work contexts that are female-dominated, but in our study the negativity is not from third party evaluators but from group members themselves.

These findings demonstrate a potential dark side to gender diversity in the workplace. As the number of women in the workplace continues to increase (projected to reach 50% by 2020; Judy & D'Amico, 1997), scholars have focused on reducing stereotype usage in evaluations of women on competence-related dimensions (e.g., Goldin & Rouse, 2000). However, such research neglects to consider how gender conceptualized at levels higher than the individual can influence evaluations of women *and* men. To the extent that gender at these levels is ignored, its effects on evaluations may remain unnoticed, leading scholars to falsely conclude that evaluations are not biased by gender. This may be the case particularly when, as in this study, the effect of gender similarly impacts men and women. Thus, our research identifies a potential "hidden" variable that may influence evaluations of individuals within task-focused settings, and consequently, those individuals' outcomes, such as their grades, promotions and raises. It also suggests that gender composition may undermine groups' functioning, to the extent that even high-performing teams with large proportions of women may not be enthusiastic about working together.

In future research it is important to distinguish between the two explanatory mechanisms we presented earlier. It may be the case that men and women evaluate each other more negatively in groups with proportionately more women because individuals are more likely to be stigmatized by association the more they interact with stigmatized others, or because the gender composition of the group colors judgments of all its members. It may also be the case that our findings are a result of both of these processes; they are not incompatible. Importantly, the two processes have very different implications for how to improve intragroup perceptions in diverse groups. The former interpersonal mechanism begs for an interpersonal solution, such as breaking down stereotypes. The
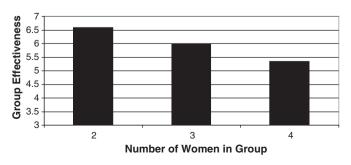

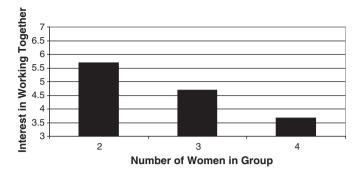**Fig. 2.** Adjusted means for ratings of group effectiveness.

Fig. 3. Adjusted means for interest in working together in the future.

latter group-level mechanism suggests a group-level solution, such as building collective efficacy to avoid "self-stereotyping" the group.

There are several other avenues for future research. It is important to systematically observe behaviors in workgroups with different proportions of women to determine whether there are differences in group process that might account for, or further elucidate, our effects. Moreover, given the temporal robustness of the group desirability effect, future research should also examine whether the experience of working in a group with proportionately more women leads individuals to avoid working in other female-dominated groups. In addition, it would be informative to explore the boundary conditions of the effects we found. For example, it remains to be seen whether they are affected by outcome dependence between group members, sex-typing of the task, or status relations within the group. Future work should also determine whether these effects generalize to other social interaction contexts in which stereotypes are applicable, such as within racially diverse task-focused groups.

Although the present research raises questions about some inadvertent by-products of gender diversity, it is not suggestive of an "anti-diversity" approach to gender in task-focused settings. Rather, it emphasizes the importance of considering how diversity might affect evaluations and sustained interest in working in gender-diverse groups in a way that was previously unexplored. In order to improve interactions between men and women in work settings, it is important to consider how gender composition of collectives (e.g., groups, departments, organizations, and professions) influences social perception processes, and in turn, shapes individuals' outcomes as well as the outcomes of those collectives.

References

Baugh, S. G., & Graen, G. B. (1997). Effects of team gender and racial composition on perceptions of team performance in cross-functional teams. *Group Organizational Management*, 22(3), 366–383.

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

Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly*, 59, 237–254.

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia: Psychology Press.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *The American Economics Review*, 90, 715–741.

Heath, C., & Staudenmayer, N. (2000). Coordination neglect: How lay theories of organizing complicate coordination in organizations. In B. M. Staw, & R. I. Sutton (Eds.), *Research in organizational behavior* (pp. 153–191). Oxford: Elsevier.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57, 657–674.

Heilman, M. E., & Wallen, A. S. (2010). Wimpy and undeserving of respect: Penalties for men's gender inconsistent success. *Journal of Experimental Social Psychology*, 46, 664–667.

Hewstone, M., & Brown, R. (1986). *Contact and conflict in intergroup encounters*. Oxford: Blackwell.

Judy, R. W., & D'Amico, C. (1997). *Workforce 2020—Work and workers in the twenty-first century*. Indianapolis, IN: Hudson Institute.

Kenny, D. A., & Livi, S. (2009). A componential analysis of leadership using the Social Relations Model. In F. J. Yammarino, & F. Dansereau (Eds.), *Multi-level issues in organizational behavior and leadership. Research in Multi-level Issues, Vol. 8*. (pp. 147–191) Bingley, UK: Emerald.

Loyd, D. L., White, J., & Kern, M. *Unkind to two of a kind: Stereotyping women with duo status in a work group*. Unpublished Manuscript.

Myaskovsky, L, Unikel, E., & Dew, M. A. (2005). Effects of gender diversity on performance and interpersonal behavior in small work groups. *Sex Roles*, 52, 645–657.

Pryor, J. B., Reeder, G. D., & Monroe, A. E. (2012). The infection of bad company: Stigma by association. *Journal of Personality and Social Psychology*, 102, 224–241.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice toward female leaders. *Journal of Experimental Social Psychology*, 48, 165–179.

West, T. V., Magee, J. C., Gullett, L., & Gordon, S. H. *The smoothing effects of perceived similarity on cross-race communication and coordination*. Unpublished manuscript.