Available online at www.sciencedirect.com
ScienceDirect
Organizational Behavior and Human Decision Processes 107 (2008) 141–160
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

Being distinctive versus being conspicuous: The effects of numeric status and sex-stereotyped tasks on individual performance in groups

Jennifer A. Chatman [a,*], Alicia D. Boisnier [b], Sandra E. Spataro [c], Cameron Anderson [a], Jennifer L. Berdahl [d]

[a] Haas School of Business, University of California, Berkeley, Berkeley, CA 94720-1900, USA
[b] Suffolk University, Sawyer Business School, 8 Ashburton Place, 6th Fl., Boston, MA 02108, USA
[c] Johnson Graduate School of Management, 365 Sage Hall Ithaca, NY 14853, USA
[d] Rotman School of Management, University of Toronto, 105 St., George Street Toronto, ON M5R 3E6, Canada

Received 9 May 2007
Available online 1 April 2008

Accepted by Dave Harrison

Abstract

Being in the numeric minority (e.g., being a solo woman in a group of men) influences how well a person performs within a work group. But being the solo member is only one way in which people can be atypical in a group; a person can also represent a social or demographic category that has not typically been associated with the task that the group is working on. Using a design with four categories of group composition (minority, balanced, majority, homogeneous) and two categories of tasks (sex-typical, sex-atypical) we found that the sex composition of the group interacted with the sex typicality of the task to influence both positive deferrals by group members and individual performance in groups. But, rather than consistently reducing performance as prior research has suggested, being numerically atypical *enhanced* individual performance when the task was typical for that person's sex. Further, positive deferrals mediated between the interaction of numeric composition and task typicality in influencing individual performance suggesting that both majority group members and the solo member affect one another's performance in groups. We conclude by discussing why understanding the interplay between these two sources of stereotyping, numeric composition and task typicality, is important for understanding the social nature of individual performance in groups.

Keywords: Solo status; Sex-congruent tasks; Diversity; Relational demography; Gender; Groups; Teams

As predictions of a major influx of women into the US workforce become reality, managers' attempts to capitalize on the purported advantages of sex diversity have met with mixed success (e.g., Jackson, Joshi, & Erhardt, 2003; Lyness & Thompson, 1997). Research on the effects of sex diversity in a variety of laboratory and organizational settings has also been characterized by mixed findings in that heterogeneity appears to be beneficial in some circumstances but detrimental in others (see Mannix & Neale, 2005; Williams & O'Reilly, 1998 for reviews). Cognitive and behavioral biases that emerge in work groups can lead people to rely on immediately apparent demographic characteristics to categorize others and predict their likely behavior within work groups (e.g., Allport, 1954; Chatman & Spataro, 2005; Messick & Mackie, 1989). A person's sex, in particular, is one of the most visible, salient, and ubiquitous social categories within work organizations (e.g., Eagly & Karau, 2002).

* Corresponding author. Fax: +1 510 845 1770.
E-mail address: chatman@haas.berkeley.edu (J.A. Chatman).

0749-5978/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2008.02.006

Kanter's (1978) early theory of proportional representation, which was based on a study of men and women in business organizations, suggested that increased visibility deriving from being numerically underrepresented generated unfair performance pressures for women, requiring that they perform at a higher level than their more numerous male counterparts to be evaluated similarly. Subsequent research supported Kanter's claim by showing that the effects of minority status are more pronounced when only one or a few demographically different people are present in a situation (e.g., Neimann & Dovidio, 1998; Steele & Aronson, 1995; Taylor & Fiske, 1978; Thompson & Sekaquaptewa, 2002). For example, women who made up a smaller proportion of the student body of a law school did less well academically than did women who attended a school in which women constituted a larger proportion, even when they were still in the minority (Spangler, Gordon, & Pipkin, 1978). Further, when men or women were solo representatives of their sex, their recall process in group discussions was disrupted by social anxiety and a distracting focus on the self (Lord & Saenz, 1985).

Just as it is likely that a person's performance in a group will decrease as the proportion of people of their sex decreases, the proportional representation view, taken to its logical extreme, implies that once numeric balance is achieved such performance problems should correspondingly decline (e.g., Abrams, Thomas, & Hogg, 1990; Pichevin & Hurtig, 1996). But, research conducted both in laboratory and field settings suggests that this decline may not occur, even when those from a previously underrepresented category begin to develop substantial representation or even outnumber those who were previously over-represented (e.g., Valian, 1998; Wharton & Baron, 1987). For example Allmendinger and Hackman (1995, p. 437) reported that organizational life for both female and male symphony orchestra members took "a qualitative turn for the worse" when the proportion of women to men reached above 10%, including increased cross-group (male/female) stereotyping and conflict and tightened identity group boundaries. Further, Chatman and O'Reilly (2004) found that women were more likely to intend to leave work groups that were numerically dominated by their own sex or balanced with respect to sex than groups that were numerically dominated by men. Taken together, this suggests that considering numeric composition alone is insufficient to explain group processes involving unbalanced representation (e.g., Crocker & McGraw, 1984; Ott, 1989).

Research in this domain has also suggested that, because of their generally lower status in society, women are derogated more than men regardless of their numeric prevalence in a particular situation (e.g., Alexander & Thoits, 1985; Crocker & McGraw, 1984; Laws, 1975; Ott, 1989; Yoder, 1991; Zimmer, 1988). For example, research has routinely shown that women's performance is more debilitated by solo status than is men's when working on various oral and written examinations, which have been more commonly the purview of men (e.g., Ben-Zeev, Fein, & Inzlicht, 2005; Sekaquaptewa & Thompson, 2002; Walton & Cohen, 2003).

But, are the effects of being a minority group member so different for men and women? Though solo status and the stereotypicality of a task for the one who performs it are different phenomena, theories and empirical studies have often failed to identify them as distinct sources of influence on a person's performance in work groups (e.g., Kanter, 1978; Pazy & Oron, 2001; also see Wood, 1987 for a review). For example, research showing that men often fare better than women in solo positions has relied primarily on tasks or occupations that have historically favored or been far more typical of men than women such as business managers, military officers, and negotiators (e.g., Cohen & Swim, 1995; Karakowsky & Siegel, 1999; Kray, Galinsky, & Thompson, 2001). Further, research in this domain has often failed to use tasks that are equivalently stereotypic for men and women (e.g., Lord & Saenz, 1985; Sekaquaptewa & Thompson, 2002). Even studies that directly intend to separate the effects of numerically based solo status and stereotype threat have not used tasks that are equivalently typical for men and women. For example, Sekaquaptewa and Thompson (2003) manipulated male and female subjects' solo status and also claimed to manipulate stereotype threat. Unfortunately, their stereotype threat manipulation involved invoking a math task, a domain for which the stereotype strongly favors men, for both the low and high threat conditions. Though Sekaquaptewa and Thompson (2003) told their subjects, in what they called the "non-stereotype threat condition," that the math test had no sex differences, the stereotype threat was likely already activated as women are a broadly identified group with respect to math (Quinn & Spencer, 2001). Karakowsky and Siegel (1999) manipulated the gender orientation of tasks through topics included in a negotiation. While the topics were indeed gendered differently as confirmed by a separate sample's response to a pre-test the act of negotiating may itself be gendered (Babcock & Laschever, 2003).

We differentiate between minority status determined by the numeric prevalence of the role occupant's sex in the current role versus minority status determined by stereotypes based on comparisons between an actual role occupant and assumptions about the types of people who typically fill, or who have historically filled, that role or performed that task. We use the term "minority," or "solo" status to refer to the numeric composition in the current role and "task typicality," or "task stereotypicality" to refer to the broader historical typicality of the frequency with which one or the other sex has performed that task or filled that role (e.g., Karakowsky & Siegel, 1999).

A key challenge in conducting research differentiating between numeric status and task typicality as sources of stereotyping is that there are remarkably few jobs that are both currently and historically performed equally by men and women (Sekaquaptewa & Thompson, 2002). Thus, distinguishing between them is difficult because they are often confounded in real-world settings. That is, people filling roles or performing tasks that are atypical for members of their social category are often also in the numeric minority, such as male nurses or female CEOs (e.g., Heikes, 1991; Yoder, 1994). Another challenge has been that studies that have distinguished between sex-congruent tasks and being a solo member of a sex in a group have often relied on vignettes. While these studies suggest that observers may evaluate people in sex-atypical roles more negatively than they do those who occupy sex-typical jobs (e.g., Cherry & Deaux, 1978; Monahan, Kuhn, & Shaver, 1974), one problem with this approach is that minority status and stereotypes are inherently social phenomena, and, as such, must be investigated in group settings. Past research has also often focused on an individual's imagined stereotypes (e.g., Steele & Aronson, 1995), but actual group behavior is essential to understand when groups are likely to defer to minority members since such deferrals likely influence a focal person's performance in the group. Thus, a minority member's performance in a group setting may depend as much on her perceptions of the typicality of the task for members of her sex as on the group's behavior toward her, that is, a group must be willing to yield to her for her to contribute to the group. Prior research has focused on intrapersonal mechanisms driving individual performance in groups, such as stereotype threat, cognitive threat, or motivation; and on performance outcomes of individuals who are differentially represented in their group or whose task is consistent or inconsistent with their own gender (Karakowsky & Siegel, 1999), but not on the mechanisms underlying such outcomes, such as the process by which group members yield to one another or confer status or leadership on one another.

Thus, we lack full understanding of how numeric status and the typicality of tasks interact precluding insight about whether, for example, being in the numeric minority can actually be an advantage, rather than a constraint, when a person from a broadly underrepresented or atypical group works on certain types of tasks. This limits our ability to unravel the sources and remedies for sex stereotypes and the challenges arising from diverse groups at work. Therefore it is critical to conduct research that examines how a group's sex composition interacts with their task, specifically whether the task is typical or atypical for one sex or the other, and how these factors jointly affect each member's performance in the interacting groups. Below we develop and test hypotheses examining the joint influence of numeric status and task stereotypicality on individual's performance in groups and on the group's behavior toward individual members. We focus on groups working on tasks with demonstrable (e.g., right and wrong) answers and who believe they are being assessed on their group's overall performance. As we have already highlighted the shortcomings of focusing on numeric status alone, we turn next to a discussion of task sex typicality and then consider their interaction.

Task sex typicality

Sex differences in behavior and associated stereotypes derive from the societal division of labor between the sexes, specifically from the different norms and expectations associated with the social roles that men and women typically hold (Carli & Eagly, 1999; Eagly & Karau, 1991). Occupations, jobs, and tasks become sex-typed as "male" or "female" according to the sex of the typical incumbent (e.g., Bielby & Baron, 1986). Men and women are expected to behave consistently with their stereotyped roles, and deviations from this typical profile—women doing a "male" task or men doing a "female" task—cause sex to become salient by challenging the expectation that the job should be, or typically has been, occupied by people of that sex and not by members of the other sex (e.g., Elfenbein & O'Reilly, 2007; Kiesler, 1975).

Mismatches between sex-typed tasks and a specific person's sex affect both observers' expectations for that person's performance (e.g., Devine, 1989) as well as the target's actual performance (Steele & Aronson, 1995). To the extent that a task stereotype is valid—that is, women have been uniquely socialized to become more capable of performing certain tasks and men of others—members of the sex that is congruent with the task are more likely to be experienced and competent at that task (e.g., Dovidio, Heltman, Brown, Ellyson, & Keating, 1988; Reskin, McBrier, & Kmec, 1999). Thus, the existence of sex stratification makes it more likely that men and women will actually have relatively more expertise with different tasks (e.g., Bielby & Baron, 1986; Cassirer & Reskin, 2000).

But once an occupation, job, or task becomes sex-typed people often use the actor's sex, rather than his actual expertise, to anticipate and judge his competence (e.g., Eagly & Wood, 1991), and these expectations influence how men and women perform on various tasks. For example, asking men to perform masculine tasks or women to perform feminine tasks caused subjects to perform worse than when they were asked to perform tasks that were identified as congruent with their sex (Hargreaves, Bates, & Foot, 1985; Sanguinetti, Lee, & Nelson, 1985). Further, men emerged as leaders

more often when the task was masculine while women emerged as leaders when the task was feminine (Wentworth & Anderson, 1984). Men also displayed more power-related behavior when working on an oil change task while women displayed more power-related behavior when working on a sewing task (e.g., Dovidio et al., 1988).

These sex-typed task effects on individual performance occur because people who represent the sex-incongruent category for the task often experience stereotype threats that disrupt their intellectual functioning and preclude them from displaying their true capabilities (e.g., Steele & Aronson, 1995). For example, group members who were imputed to be experts on a task, regardless of their actual expertise, were more actively engaged in the group's activities than were those who were not imputed experts (Karakowsky & McBey, 2001). Further, equally qualified women performed substantially worse than men when they were told that a test produced sex differences (high stereotype threat) than when they were told that the test did not produce sex differences (low stereotype threat) (e.g., Spencer, Steele, & Quinn, 1999).

Taken together, existing research on numeric status and task stereotypicality suggests that a person would perform at his or her worst when: (1) there are fewer, rather than more, members of the same sex in their work group (e.g., Kanter, 1978; Neimann & Dovidio, 1998) and (2) their sex is inconsistent with the stereotype for the task (e.g., Eagly & Wood, 1991). But a critical question remains; how do these two conditions interact? By underemphasizing their distinct influence on individual performance, past research has, by default, treated numeric status and task typicality as though they combine additively (Cohen & Swim, 1995). We suggest instead that they may interact in a more complex way, which we specify below.

How task sex typicality influences the relationship between numeric status and individual performance in groups

Prior research has shown, convincingly, that solo men working in groups of women gain advantages such as attaining power and boosting evaluations of their performance (e.g., Crocker & McGraw, 1984; Fairhurst & Snavely, 1983; Floge & Merrill, 1986; Fottler, 1976; Heikes, 1991; Kadushin, 1976; Ott, 1989; Williams, 1992; Yoder & Sinnett, 1985). Underlying this research is a view that these advantages derive from men's higher societal status (e.g., Eagly & Karau, 2002; Inzlicht & Ben-Zeev, 2000, 2003). In contrast, we suggest that past research has not typically conducted fair tests that would allow us to understand the joint impact of solo status and sex task congruence on men and women. Thus, we suggest that both solo men and women will experience a boost in performance when the task is typical for their sex and a corresponding decline in performance when the task is atypical for their sex. Below we consider how a person is perceived and treated by others in the group and their own motivation and ability to perform are jointly influenced by being a solo member of their sex and the stereotypicality of the group's task for their sex.

The impact of numeric status and task sex typicality on positive deferrals

We suggest that task typicality and numeric composition determine whether solo status will be detrimental or advantageous to how other members perceive and treat a focal member. Specifically, other group members will place lower value on solo men or women working on a task that is stereotyped as incongruent with their sex because those members assume that the solo member has little competence on the task. As a result of these lowered expectations, group members will be less motivated to involve the task-atypical solo member in the task since they do not perceive that including him or her will accomplish the task more successfully. Further, the relative discomfort of working with dissimilar rather than similar others can exacerbate this tendency to neglect the solo member (e.g., McPherson, Smith-Lovin, & Cook, 2001). Indeed, research on social networks has shown that minority members are excluded by majority members (e.g., Mehra, Kilduff, & Brass, 2001). As a result, solo members whose sex is incongruent with the task are more likely to be isolated and uncomfortable about engaging in the task (e.g., Konrad, Winter, & Gutek, 1992) precluding group members from gathering accurate information about the solo member's true capabilities on the task (e.g., Flynn, Chatman, & Spataro, 2001). In contrast, research suggests there may be some conditions under which a minority group member might be more influential. In their study of corporate board members, Westphal and Milton (2000) found that the influence of gender minorities increased when they had previous board experience in either a minority or majority role, when majority members had previous board experience in a minority role, or when minority members had direct or indirect ties through membership on other boards with majority board members. Building on these findings, we additionally posit that when a solo man or woman works in a group focusing on a sex-congruent task, he or she may receive more positive attention from other group members.

Indeed, such numerically distinct perceived "experts" might receive even more attention, support, and deference from their colleagues than would those who are in the numeric majority (Sekaquaptewa & Thompson,

2002; Taylor & Fiske, 1978). In essence, their solo status heightens the salience of their unique expertise. As Brewer's (1991) theory of optimal distinctiveness suggests, the harmful effects of being a solo member dissipate when the individual is high status. In the current context, being perceived as the only expert on the task, by virtue of its sex congruence, is likely associated with higher status in the group (e.g., Berger, Webster, Ridgeway, & Rosenholtz, 1986), and, in contrast to the solo for whom the task is atypical, may even induce the solo member working on a typical task for his or her sex to become more central to the group. A key behavioral indicator of the extent to which a group member is valued is positive deferrals (e.g., Driskell, Olmstead, & Salas, 1993), which we define as statements by group members that explicitly highlight a focal individual's relative expertise or invite her to offer her view. Group members should be more likely to defer to the member of their group whom they believe will bring them success, namely the solo "expert" compared to the solo member whose sex is incongruent with the task (Berger & Zelditch, 1985).

It is also important to consider how sex stereotypicality of the task influences members of other numeric status in groups. Drawing on our logic above, we expect that the uniqueness of the solo member, that he or she is the only one of the other sex, causes other members to devote an inordinate amount of attention to him or her. In the case of a solo member working on a sex-typical task, this results in the highest level of positive deferrals compared to members of all other numeric combinations—that is, members of groups that are balanced with respect to sex, members of the majority sex, and members of homogenous groups—one sex or the other—all working on either type of task. Thus, it is the uniqueness of the solo member's perceived expertise on a sex-congruent task that increases other members' positive deferrals to him or her. In contrast, using the same logic, we expect that solo members working on sex-atypical tasks will receive the lowest level of positive deferrals of all combinations of numeric status and sex stereotypicality. Taken together, the preceding arguments suggest two hypotheses regarding the solo member and other members of variously composed groups, specifically,

Hypothesis 1A: Controlling for a person's actual expertise on the task, solo members of their sex who are working on a task that is typical for their sex will be positively deferred to in a group significantly more often than will solo members who are working on a task that is atypical for their sex.

Hypothesis 1B: Controlling for a person's actual expertise on the task, sex stereotypicality of the task will moderate the relationship between numeric status and the extent to which group members positively defer to that person such that the effect of being a solo member (compared to being any other combination of numeric status, such as balanced, majority, or homogeneous) while working on a task that is typical for one's sex on positive deferrals received will be more positive than the effect of being a solo member while working on a task that is atypical for one's sex.

The impact of numeric status and task sex typicality on individual performance in groups

In addition to provoking intensified attention among group members, numeric status also likely interacts with the sex congruence of the task to influence solo members' actual behavior and effectiveness in a group. In particular, solo members of a group whose sex matches the task should be less likely to engage in social loafing compared to members whose sex is congruent with the task but are in the numeric majority, or to those solo members in sex-incongruent task situations. Being one of multiple group members representing the sex that typically undertakes the task may result in a diffused sense of accountability and lowered performance expectations resulting in lower effort and individual performance, a notion long supported by theories of free-riding and social loafing (e.g., Ingham, Levinger, Graves, & Peckham, 1974; Kerr & Brunn, 1983; Latane, Williams, & Harkins, 1979). For example, a person's contributions to a group task tends to decline as the number of members increases (e.g., Hart, Bridgett, & Karau, 2001; Karau & Williams, 1993). We suggest that this will be more likely among majority same-sex members, for whom the task is incongruent with their sex because those members will be more likely to free ride on others whom they perceive to be able to perform the task competently, in this case implicating those solo members for whom the task is sex-congruent.

The notion that sex-typical members would perform better when they are in the numeric minority seems, on it face, to contradict various lines of research and theory. In particular, Lau and Murnighan's (2005) theory of group faultlines suggests that minority members' performance should suffer when minority/majority distinctions in groups are related to the group's task. We believe this is only an apparent contradiction, however. As Lau and Murnighan (2005) point out, when groups are under pressure to work together toward a common goal, they are likely to reduce the negative valence of their demographic differences and do what they believe is needed to succeed, which is rely on the solo member for whom the task is typical.

In contrast, stereotype threat research suggests that minority members should perform especially poorly when they are working on a task that is not stereotypically linked to their sex (Steele & Aronson, 1995). Being

a solo member of a group is often uncomfortable and can reduce self-esteem (Brewer, 1991). Further, minority members fall prey to cognitive distractions, such as arousal, when their minority status is salient (Ben-Zeev et al., 2005). But stereotype threats should only harm one's performance when the stereotype suggests that one's sex does poorly on the task, not when it suggests that one's sex performs well (Fuegen & Biernat, 2002; Steele & Aronson, 1995). Thus, these negative forces should not just disappear, but reverse when the minority member is working on a task that is positively linked to their sex.

Further, using the same logic as above, we expect that solo members should exhibit the most extreme levels of performance, with solos working on sex-congruent tasks showing the highest level of performance and solos working on sex-incongruent tasks displaying the lowest level of performance compared to all other combinations, that is, members of groups that are balanced with respect to sex, members of the majority sex working on either type of task, and members of homogenous groups—one sex or the other all working on either a sex-typical or -atypical task. Taken together our arguments suggest the following hypotheses,

Hypothesis 2A: Controlling for a person's actual expertise on the task, solo members of their sex who are working on a task that is typical for their sex will perform significantly better in the group than will solo members who are working on a task that is atypical for their sex.

Hypothesis 2B: Controlling for a person's actual expertise on the task, sex stereotypicality of the task will moderate the relationship between numeric status and individual performance in the group such that the effect of being a solo member (compared to being any other combination of numeric status, such as balanced, majority, or homogeneous) while working on a task that is typical for one's sex on individual performance will be more positive than the effect of being a solo member while working on a task that is atypical for one's sex.

How positive deferrals mediate the relationship between numeric status and individual performance in groups

One remaining question involves identifying the mechanism underlying the relationship between solo status, task sex typicality, and the solo members' performance in groups. Above we presented arguments for how the interaction between solo status and task sex typicality influences positive deferrals and individual performance in groups. Though we suggested that the interaction directly influences the solo individual to perform differently in a group depending on the typicality of the task for their sex, it is also possible that positive deferrals constitute the mechanism underlying this relationship such that the presence of positive deferrals enhances performance among solo members working on a sex-typical task. That is, as we argued above, unless group members defer to a solo member of the other sex, his or her performance will be constrained in the group. Thus, social facilitation effects (e.g., Zajonc, 1965) occur from the increased attention the solo member receives from others in the group who are interested in accomplishing the task at the highest level. This increased positive attention enables the solo member to exhibit his or her task relevant competence. Therefore, we predict that,

Hypothesis 3: Controlling for a person's individual expertise on the task, positive deferrals will mediate the interaction of solo status with task sex-stereotypicality such that the direct effect of the interaction of task sex congruence and solo status on performance will weaken after positive deferrals are considered.

Method

Sample

Three hundred sixty students working in 90 four-person groups enrolled in undergraduate psychology courses at a major West Coast university participated in this experiment as part of their course requirements.[1] Forty-nine percent of our subjects were male, subjects' average age was 19.76 years ($SD = 2.79$), 51% were Asian or Asian-American, 37% were Caucasian, 7% were of Latino/Hispanic descent, 3% were African-American, and 2% identified themselves as "other." This distribution closely resembled the University's student population. Subjects were assigned to experimental sessions through a combination of volunteer sign-ups and phone calls inviting them to participate.

Research design and procedure

We conducted a laboratory experiment to test our hypotheses, running a 4 (homogeneous member, majority member, balanced member, and minority member)-by-2 (sex-typical task, sex-atypical task) design. Each

[1] An additional 162 subjects were excluded from the sample described above due to our intentional over-scheduling of participants. They were randomly chosen and not assigned to experimental groups and were instead asked to complete a survey that was unrelated to this study and then dismissed with full credit for participating. Analyses showed no difference on all demographic indicators, between participants who were and were not included in four-person groups.

session of the experiment took approximately 1.5 h and was divided into three phases. In the first phase, participants were told that the study was designed to test their ability to solve either math or verbal problems and that they would be asked to work individually and then in a small group to solve such problems. Participants were also informed that the group they worked with would have the opportunity to earn $100 based on its performance. Participants then worked individually behind a screen set up at each seat to provide privacy while working on the task. Participants worked for 10 min on a set of standardized test problems randomly selected from previous Graduate Management Aptitude Tests (GMAT), used primarily for selection into graduate schools of business (e.g., Hecht & Schraeder, 1986).

Following past research that uses this test or other similar standardized tests (e.g., Aronson et al., 1999), we chose the GMAT to insure that the problems were challenging for our undergraduate participants and because the test is naturally divided into sex-typed tasks in that it contains a significant math section and a significant verbal section. Research has routinely shown that math and verbal tasks are sex typed (Walton & Cohen, 2003). Mathematics and disciplines drawing heavily on it (e.g., computer science) have historically been, and continue to be, largely dominated by male students and faculty (e.g., Kirkman, Maxwell, & Priestley, 2003). For example, at the University where this study took place, 78% of general engineering majors, 75% of physics majors, and 83% of computer science majors were male. Disciplines with a majority of women (e.g., English) draw more heavily on verbal than on math skills. At the University where this study took place, 68% of English majors and 74% of communication majors were female. Participants in our study were, therefore, randomly assigned to work on either math or verbal GMAT questions.

In the second phase of the experiment, participants were randomly assigned, within sex, to a group of four. The sex composition of each group was either minority/majority (one member of one sex and three from the other sex), balanced (two members of each sex), or homogeneous (all members of one sex or the other). The task type matched the task each subject had worked individually on in the first phase of the experiment (math or verbal, which depending on the subject's sex, corresponded exactly to whether it was typical or atypical for their sex). The experimenter called the names of participants assigned to each group in random order and instructed them to follow one of the other experimenters to a separate room where they would be videotaped working on another set of GMAT questions as a group. Teams were instructed to work as a group rather than individually in this second phase of the experiment. We provided each group with only one set of question-and-answer materials. Each group received either 91 verbal problems or 142 math problems,[2] depending on their task condition, which were bound together in a booklet of overhead transparencies. Groups were given 40 min to work on the GMAT questions in a room equipped with an overhead projector and a table with four chairs.

Subjects recorded their group's answer to each question on a single scantron form. Experimenters explicitly instructed participants to work together as a group during this phase and reiterated the $100 prize incentive for the "best" group. Participants were told:

At this time, we ask that you work as a group to solve the [math/verbal] problems. Like in the first part of the study, you may only use the materials provided to work on the problems... This time, you should work together to solve the problems. There is a prize of $100 for the best group. Please note that there are more problems here than most groups can finish in the given amount of time. Try to do as many as you can. You will be penalized for wrong answers but not as much as you will be rewarded for correct answers. The prize will go to the team that both finishes the most problems and works together the most.

Despite our instructions to work together on the task, four groups did not follow these directions and instead took apart the bound overhead packet and worked individually on the group task. Because these members did not speak or interact during the time that they worked on the actual problems, it was impossible to code their positive deferral behaviors and their performance for use in our analyses. Therefore, we excluded these four groups from our analyses, reducing the valid sample size from 90 groups to 86 groups ($n = 344$). These four groups were spread across conditions (1 all female, 1 balanced, 2 majority male), and no significant differences existed between these individuals and the full sample on all demographic variables.

In the third phase of the experiment, participants individually completed a post-experiment questionnaire. Participants from both groups were instructed to return to the room in which they had originally completed the individual task. Each privacy screen was labeled with participants' names (in random order) and an identification letter (e.g., W, X, Y or Z) so that participants could identify one another while completing the questionnaire. Participants were asked to sit in the seat corresponding to their name and to respond to the questionnaire.

[2] We developed the number of problems assigned based on a proportion of the amount of time that would have been allotted by the Graduate Management Admissions Council to GMAT test-takers to complete the exam under normal test-taking conditions (e.g., Hecht & Schraeder, 1986).

Independent variables

Numeric status

We created groups with four types of sex composition to manipulate numeric status. A total of 24 minority-female groups (one female and three males) and 24 minority-male groups (one male and three females) constituted minority and majority conditions; 23 balanced groups (two females and two males) constituted the balanced condition; and 15 groups (6 all male and 9 all female) constituted the homogeneous condition. Thus, a subject's numeric status varied according to whether they were a minority ($n = 48$), balanced ($n = 92$), majority ($n = 144$), or homogeneous ($n = 60$) member of their group.

Sex-stereotyped tasks

We manipulated the sex-stereotype of the task by dedicating each experimental session to either GMAT math or verbal problem sets. Although research has routinely shown that math and verbal tasks are sex typed (Walton & Cohen, 2003), we wanted to insure that sex based stereotypes existed for these tasks in the population we studied. Therefore, we conducted a pre-test, following previous research (e.g., Karakowsky & Siegel, 1999), to determine whether the population of students from which we drew our sample held the expected stereotypes relevant to this study—that men more typically undertake math-related tasks and have superior math skills than do women and that women more typically undertake and have superior verbal skills than do men. We distributed surveys to an independent sample of 47 students (40% women) who attended the same University and were identical in demographic make-up to our subjects, asking these students to report their own opinions as well as the "average student's opinion at this University." We considered this to be a check on our manipulation of task sex typicality as the samples for the pre-test and experiment was drawn from exactly the same population. Further, we were concerned that asking such questions of our experimental subjects before they participated in the individual or group part of the experiment would artificially heighten their awareness of the sex composition of their group leading to demand effects, and asking these questions after the experiment was over on the post-experiment survey would confound their responses with the condition to which they had been randomly assigned.

Pre-test respondents were asked to indicate "whether men or women perform better in the following areas" on a scale from 1 to 5 (1 = women perform much better, 3 = neither men nor women perform better, and 5 = men perform much better): (a) writing and grammar tasks, (b) math tasks, (c) the Scholastic Aptitude Test (SAT), (d) the verbal portion of the SAT, and (e) the math portion of the SAT. Respondents also indicated whether majors dominated by males or females required more: (f) writing, grammar, and verbal skills, and (g) more math skills (1 = majors dominated by females, 2 = neither male nor female dominated majors, 3 = majors dominated by males).

Pre-test respondents displayed the stereotypes we expected on every item, and indicated that the "average student at this University" would too. Responses on each item were significantly different from neutrality in the direction of traditional sex stereotypes (mean across items = $|t| \geqslant 2.12$, $p \leqslant .05$). It is important to recall that our pre-test and study samples did not overlap. Therefore, the pre-test provided an uncontaminated assessment of the extent to which subjects drawn from this population held the expected sex stereotypes regarding math and verbal tasks.

We also wanted to insure that the GMAT task was challenging enough to avoid ceiling effects. To do this, we selected GMAT questions that spanned the levels of difficulty one would see on the actual test. We also chose the GMAT (rather than the SAT, a similar standardized test used in past research and for admissions to college) because it is a test for admissions to graduate school, while our subjects were undergraduates. Further, judging from participants' performance, the problems appeared to be quite difficult to solve and there were more problems presented than any of the groups could solve completely and correctly. The groups answered between 0% and 73% of the questions correctly ($M_{\text{overall}} = .29$, $SD = .14$; $M_{\text{verbal}} = 52.99\%$, $SD = 19.84\%$, $M_{\text{math}} = 39.82\%$, $SD = 17.71\%$).

Dependent variables

Using a video tape of each group's problem-solving session, independent judges, who were unaware of the study's design and hypotheses, coded the group members' behaviors as outlined below.

Positive deferrals

We coded behavior in the work group meetings shown on the videotapes and created a variable reflecting the extent to which each subject was positively deferred to by his or her colleagues. Using a form that we created specifically for this purpose, four independent coders who were blind to the study design and hypotheses separately tallied the number of times each group member was positively deferred to by other group members. Examples included asking, "What do you think, [name]?" or stating, "Yeah, you're right, [name], that's good." Two independent coders were assigned a common set of videotapes and another two were assigned another common set of videotapes (with overlap between the pairs to assess reliability). The coders agreed substantially on their tallies, producing inter-rater reliability correlations of $r = .87$, thus we averaged

the raters' tallies for each individual, for a final value of positive deferrals received for each individual ($M = 1.87$; $SD = 1.79$).

Individual performance in the group

We operationalized individual performance in the group as active, verbal contribution of correct answers by an individual member of the group during the group task. Again, four independent coders who were unaware of the study design or hypotheses used videotapes of the groups' problem-solving sessions to code each group member's contribution. Each coder watched each experimental session separately and in random order. Coders were instructed to tally the number of times each subject verbally expressed: (a) the correct answer to the question and (b) the answer the group reported on their scantron answering form. Regardless of the number of times a subject mentioned the correct or reported answer, each subject received a maximum of one point per response type per problem completed by the group. Coders were instructed to count an answer only if the participants actually stated an answer choice (A, B, C, D, or E) of their own accord. If the participants simply agreed with or repeated another group member's answer (e.g., "Yeah" or "That's right") without stating an answer choice, it was not counted. Therefore, "I think the answer is B" was counted, but "Yeah, I think that's the right answer," or "I agree" were not. We made this distinction in order to avoid attributing genuine contributions to members who merely "tagged on" to others' answers to appear to be contributing when they may not have known the answer.

Again, two independent coders were assigned a common set of videotapes and the other two were assigned another common set of videotapes (with overlap to assess reliability). Coders agreed substantially on their tallies of both variables for "gave the correct answer" ($r = .89$), and "gave the answer the group reported" ($r = .87$). The two variables also had high scale reliability (coefficient alpha = .97), so we averaged the coders' tallies for each subject and then averaged across the two measures of performance to form one performance score for each subject. We then divided subjects' performance scores by the total number of answers their group recorded on their answer sheet, yielding a performance score representing the percent of items for which each subject provided answers out of the total number of items for which the group gave an answer on its group-reporting scantron form.

Finally, subjects working on verbal problems performed consistently higher than did those working on math problems in terms of the percentage of problems they answered correctly ($M_{\text{verbal}} = 52.99\%$, $SD = 19.84\%$, $M_{\text{math}} = 39.82\%$, $SD = 17.71\%$; $t = 3.80$, $p < .01$). We therefore standardized the performance ratings by centering them around the mean verbal and math scores. After standardizing, the mean level of individual performance across standardized verbal and math conditions was.06 ($SD = .99$).

This performance measure is consistent with our operational definition of individual performance in the group in this context, which we define as actively contributing to the correct answer. Videotaped observations are limited, however, in their ability to capture group members' subjective perceptions of one another's contributions expressed by other behaviors such as participating in group discussion, making an effort to contribute, and contributing in more passive ways such as facilitating cooperation and reducing conflict within the group. Therefore, we assessed the validity of the videotaped performance measure by correlating it with team members' ratings of one another's performance during the group task. To create these *peer performance ratings*, we averaged the ratings individuals received from their group members on each of 13 performance-related items presented on the post-task survey. Items began: "During the group task, to what extent did each member of the group..." and ended with phrases like "participate," "make valuable contributions," and "demonstrate high levels of effort." Each group member was rated on a scale of 1 ("not at all") to 7 ("a great deal"). A factor analysis of these 13 items showed that they loaded onto a single factor (item loadings ranged from .50 to .79). The coefficient alpha ($\alpha = .94$) showed that the scale was reliable. We therefore combined the 13 items to create one scale measure of peer-rated performance ($M = 5.25$, $SD = .45$). As a final step we standardized peer-rated performance within task condition (math or verbal) since average performance ratings, again, differed by task. We found a positive correlation between peer-rated performance and coders' ratings of individual performance in the group ($r = .43$, $p < .01$), suggesting that the videotaped performance measure substantially reflected how people evaluated each others' performance during the task.

Control variables

Individual expertise

We included an indicator of expertise as a covariate in all analyses. We operationalized individual expertise as the percentage of problems presented that each subject answered correctly on the individual task during the first phase of the experiment ($M = 46.63$, $SD = 20.07$). Like individual performance, we standardized expertise around task means to reduce the discrepancy between math and verbal performance on the individual task ($M = -.05$, $SD = 1.01$).

Interdependence

The videotaped observations revealed variation in the extent to which individuals worked together during the

Table 1
Descriptive statistics and intercorrelations among study variables

Variables	Mean[a]	*SD*	1	2	3	4	5	6	7	8
Controls										
1. Interdependence (0 = high, 1 = moderate)	.02	.15	—	—	—	—	—	—	—	—
2. Expertise[b]	−.05	1.01	−12*	—	—	—	—	—	—	—
Independent										
Numeric status										
3. Minority = 1	.14	.35	.06	.03	—	—	—	—	—	—
4. Balanced = 1	.27	.44	.09	−.06	—	—	—	—	—	—
5. Majority = 1	.42	.49	.13*	.14*	—	—	—	—	—	—
6. Homogeneous = 1	.17	.38	.34**	.15*	—	—	—	—	—	—
7. Task sex-stereotypicality (0 = typical, 1 = atypical)	.49	.50	.16*	−.14**	.03	.02	−.02	−.02	—	—
Dependent										
8. Positive deferrals	1.87	1.79	.14*	.24**	.12*	−.02	.01	−.10	−.01	—
9. Individual performance[b]	.06	.99	−.19**	.41**	.00	−.02	−.03	.06	−.09	.30**

[a] Entries in the mean column are proportions for categorical variables.
[b] Standardized around task means.
* $p, < .05$, two-tailed tests.
** $p < .01$, two-tailed tests.

group task. Four blind, independent coders observed that for some or all of the duration of the group task, a total of six groups (1 all female, 1 all male, 1 majority female, 1 balanced, 2 majority male) either worked in pairs rather than as a whole group or worked in a group of three while one member worked alone. Coders agreed completely on which groups displayed less interdependent group work behavior ($r = 1.00$). Unlike the four groups that exhibited no interaction at all discussed above, these six groups did exhibit interaction, permitting us to code positive deferrals and individual performance for each member. We therefore, created a dummy variable for interdependence in which we assigned 0 to groups that worked together as a four-some throughout the duration of the task, and 1 to those groups that worked in some other combination but still interacted, and included this variable as a covariate in all of our analyses.

Analyses

We used ANCOVA to examine the interactive effects of proportional representation (minority, balanced, majority, homogeneous) and task sex-stereotypicality (sex-typical task, sex-atypical task) on individual contribution and positive deferrals. We included individual expertise and interdependence as covariates. We conducted pairwise comparisons to identify the nature and form of the significant interactions. Table 1 shows the descriptive statistics and intercorrelations among the study variables.

Data within groups can violate assumptions of independence. According to Kenny and LaVoie (1985, p. 344), if such non-independence exists, it is best to analyze data at the group level. If data within groups is not correlated, individual level data can be meaningfully interpreted. We calculated the intra-class correlation for each dependent variable (Kenny & LaVoie, 1985) and it was .01 for positive deferrals and .00 for individual performance in groups, indicating that within-group data were not correlated. Therefore, we conducted our analyses at the individual level.

Results

In hypothesis 1A, we predicted that minority members working on tasks that were typical for their sex would receive more positive deferrals from their teammates than would minority members working on atypical tasks. Pairwise comparisons (see Table 2) revealed that solo subjects working on sex-stereotypical tasks received significantly more positive deferrals from their team mates ($M = 3.08$, $SD = 3.11$) than did solo members working on atypical tasks [$M = 1.75$, $SD = 1.24$; $F(1,48) = 4.29$, $p < .05$].[3]

In hypothesis 1B, we predicted an interaction such that the effect of being a solo member (compared to being any other combination of numeric status, such as balanced, majority, or homogeneous) while working on a task that was typical for one's sex on positive deferrals received would be more positive than the effect of being a solo member while working on a task that was atypical for one's sex. As predicted, a significant interac-

[3] We report simple (raw) means in the text. Estimated marginal means for the full equations are reported in Table 2.

Table 2
Estimated marginal means, confidence intervals, and comparison of means for each experimental condition

Condition	Positive deferral behaviors received by individuals			Individual performance[a] in groups		
	Mean	95% Confidence interval	Comparison of means[b]	Mean	95% Confidence interval	Comparison of means
Minority						
1. Typical	3.06 (.35)	2.37, 3.75	1 vs 2*	.46 (.18)	.10, .81	1 vs 2**
			1 vs 3+			1 vs 3*
			1 vs 4*			1 vs 4
			1 vs 5**			1 vs 5**
			1 vs 6*			1 vs 6*
			1 vs 7**			1 vs 7
			1 vs 8			1 vs 8
2. Atypical	1.80 (.34)	1.13, 2.47	2 vs 3	−.43 (.17)	−.77, −.09	2 vs 3+
			2 vs 4			2 vs 4*
			2 vs 5			2 vs 5
			2 vs 6			2 vs 6+
			2 vs 7+			2 vs 7**
			2 vs 8			2 vs 8**
Balanced						
3. Typical	1.93 (.25)	1.44, 2.42	3 vs 4	−.05 (.13)	−.30, .21	3 vs 4
			3 vs 5			3 vs 5
			3 vs 6			3 vs 6
			3 vs 7*			3 vs 7
			3 vs 8			3 vs 8**
4. Atypical	1.92 (.25)	1.43, 2.41	4 vs 5	.11 (.13)	−.14, .36	4 vs 5
			4 vs 6			4 vs 6
			4 vs 7*			4 vs 7
			4 vs 8			4 vs 8**
Majority						
5. Typical	1.78 (.19)	1.39, 2.17	5 vs 6	−.11 (.10)	−.31, .09	5 vs 6
			5 vs 7+			5 vs 7*
			5 vs 8			5 vs 8**
6. Atypical	1.95 (.21)	1.54, 2.35	6 vs 7*	−.04 (.11)	−.25, .17	6 vs 7+
			6 vs 8			6 vs 8**
Homogeneous						
7. Typical	1.18 (.30)	0.60, 1.77	7 vs 8	.28 (.15)	−.02, .58	7 vs 8+
8. Atypical	1.65 (.37)	0.91, 2.39		.76 (.19)	.38, 1.14	

Note. Estimated marginal means are for full ANCOVA equation. Values enclosed in parentheses represent standard errors.
[a] Standardized around task means.
[b] Comparisons using univariate ANCOVAs.
* $=p < .05$.
** $=p < .01$.

tion of numeric status and sex-stereotyped tasks on being positively deferred to by others emerged [$F(3, 344) = 2.63$; $p < .05$; see Table 3 and Fig. 1]. We conducted two types of contrasts to explore the form of the interaction. First, we conducted Univariate ANCOVA's comparing positive deferrals received by solo typical or atypical subjects, respectively, to all other conditions aggregated together. Solo members working on typical tasks received significantly more positive deferrals ($M = 3.08$, $SD = 3.11$) than did members in all other conditions [$M = 1.79$, $SD = 1.63$; $F(1, 344) = 12.13$, $p < .01$]. Second, we conducted more detailed condition-to-condition post hoc comparisons. The results of the specific contrasts are shown in Table 2 and Fig. 1. We found that solo members working on tasks that were typical for their sex received significantly more positive deferrals than did all other subjects except for those in the homogeneous groups working on a task that was atypical for their sex [$M = 1.94$, $SD = 1.93$; $F(1, 51) = 1.35$, n.s.].

Solos working on an atypical task compared to all other conditions received, however, no fewer positive deferrals ($M = 1.75$, $SD = 1.24$) than did other subjects [$M = 1.88$, $SD = 1.83$; $F(1, 344) = .06$, n.s.] (except, of course, than solos working on typical tasks). And, in our more detailed paired comparisons, we found that solos working on the atypical task received a marginally greater number of positive deferrals, rather than fewer as we predicted than did those in one other condition, the homogeneous groups working on typical tasks [$M = 1.12$, $SD = 1.22$; $F(1, 54) = 3.82$, $p < .10$]. Thus, hypothesis 1B was supported for solos working on a

Table 3
Analysis of covariance for main outcome variables

Source	Positive deferrals			Individual performance		
	df	*F*	Power (d^2)	*df*	*F*	Power (d^2)
Covariates						
Interdependence	1	9.95**	.89	1	20.14**	.99
Expertise[a]	1	23.77**	1.00	1	75.01**	1.00
Main effects						
Numeric status (NS)[b]	3	2.92*	.69	3	5.77**	.95
Task sex-stereotypicality (TS)[c]	1	.60	.12	1	.19	.07
Interaction effects						
NS × TS	3	2.63*	.64	3	5.83**	.95
Error	334	(2.86)		334	(.75)	
Model R^2	.13			.26		
Model *N*	344			344		

Note. Values enclosed in parentheses represent mean square errors.
[a] Standardized around task means.
[b] NS = majority, balanced, minority, or homogeneous member.
[c] TS = typical or atypical.
* $p < .05$, two-tailed tests.
** $p < .01$, two-tailed tests.

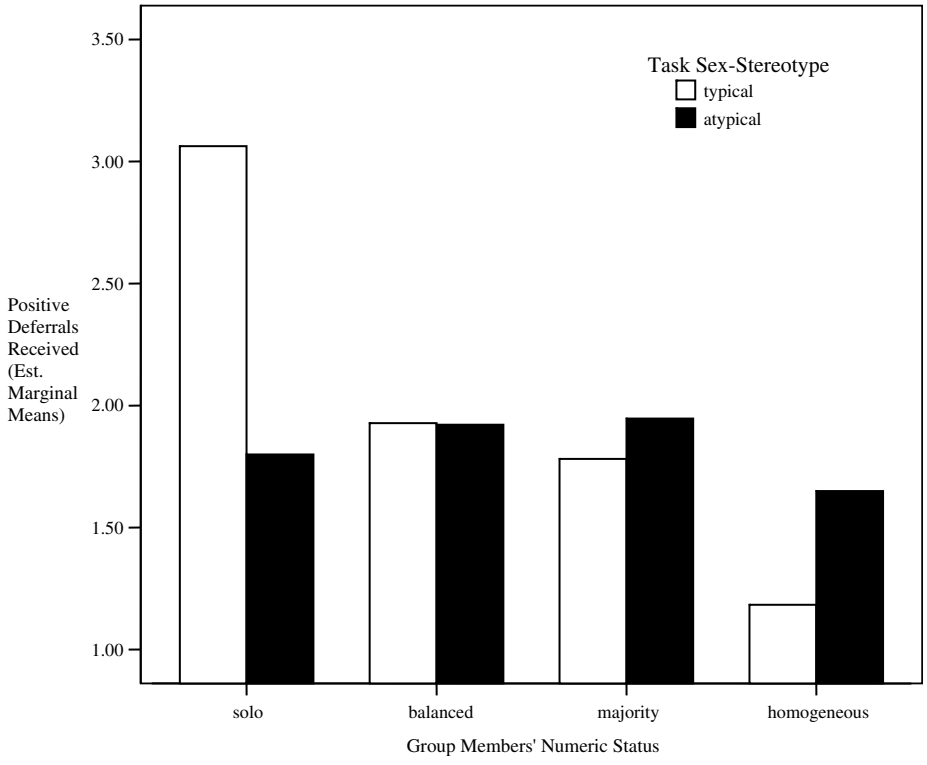

Fig. 1. Interaction effects of numeric distinctiveness and task sex-stereotype on number of positive deferrals received.

sex-typical task, but not for solos working on a sex-atypical task.

In hypothesis 2A we predicted that solo members of their group would perform significantly better when working on a sex-typical than on a sex-atypical task. Pairwise contrasts (see Table 2 and Fig. 2) showed that, above and beyond their individual expertise on the task assessed in the first phase of the experiment, minority members working on a sex-stereotypical task in their group performed significantly better ($M = .56$, $SD = .97$) than did minority members working on a task that was not typical for their sex [$M = -.39$, $SD = 1.00$; $F(1, 48) = 11.25$, $p < .01$], supporting hypothesis 2A.

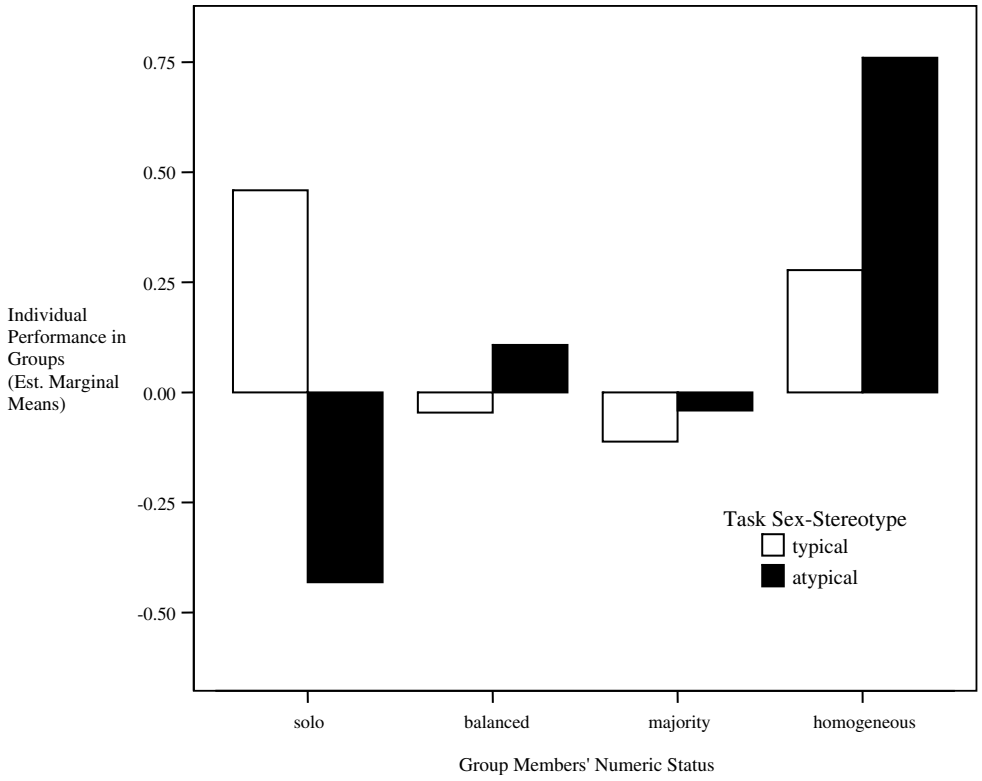

Fig. 2. Interaction effects of numeric distinctiveness and task sex-stereotype on individual performance.

In hypothesis 2B we predicted an interaction such that the effect of being a solo member (compared to being any other combination of numeric status, such as balanced, majority, or homogeneous) while working on a task that was typical for one's sex on individual performance would be more positive than the effect of being a solo member while working on a task that was atypical for one's sex. The interaction was significant, as predicted: $F(3,344) = 5.83$; $p < .01$ (see Table 3). Again, we conducted the two types of tests that we used for positive deferrals, as described above, to examine the form of the interaction. Table 2 and Fig. 2 show the results of these contrasts. Solos working on the typical task performed better ($M = .56$, $SD = .97$) than did subjects in all other conditions combined [$M = .03$, $SD = .98$; $F(1,344) = 5.43$, $p < .05$. The more detailed paired comparisons showed that solo members working on a sex-stereotypical task performed significantly better than did members of balanced groups working on a typical task ($M = -.09$, $SD = .95$; $F(1,69) = 5.58$, $p < .05$), majority members working on sex-stereotypical tasks ($M = .09$, $SD = .95$; $F(1,99) = 7.61$, $p < .01$), and than majority members working on tasks that were atypical for their sex ($M = -.04$, $SD = .88$; $F(1,91) = 6.51$, $p < .05$). Solo members working on sex-stereotypical tasks performed no better, however, than did members of balanced groups working on atypical tasks [$M = .15$, $SD = 1.00$; $F(1,69) = 2.55$, n.s.], or than homogeneous groups working on either the atypical task [$M = .07$, $SD = 1.21$; $F(1,51) = .92$, n.s.], or on the sex-stereotypical task ($M = .30$, $SD = 1.05$; $F(1,55) = .49$, n.s.). Thus, hypothesis 2B was only partially supported for solos working on the sex-typical task.

Minority members working on a task that was atypical for their sex performed significantly worse than others according to both of our tests. In the overall Univariate ANCOVA, minority members working on a task that was atypical for their sex performed significantly worse ($M = -.39$, $SD = 1.00$) compared to all other member types combined [$M = .10$, $SD = .98$; $F(1,344) = 7.85$, $p < .01$]. And in the paired comparisons, we found that solos working on the atypical task performed significantly worse than did those in the balanced atypical [$M = .15$, $SD = 1.00$; $F(1,71) = 6.03$, $p < .05$], homogeneous typical [$M = .30$, $SD = 1.05$; $F(1,57) = 7.54$, $p < .01$] and homogeneous atypical [$M = .07$, $SD = 1.21$; $F(1,53) = 13.56$, $p < .01$] conditions; and marginally worse than those in the balanced typical [$M = -.09$, $SD = .91$; $F(1,71) = 3.03$, $p < .10$] and majority atypical [$M = -.04$, $SD = .88$; $F(1,93) = 3.49$, $p < .10$] conditions; but no worse than those in the majority typical condition [$M = .10$, $SD = .95$; $F(1,101) = 2.47$, n.s. Taken together, these results offer reasonably strong support for hypothesis 2B with respect to solos working on the sex-atypical task.

Finally, to test hypothesis 3 we examined the interaction of solo status and task stereotypicality compared to all other numeric status conditions combined (i.e., majority, balanced, and homogeneous group members) with task sex-stereotypicality to determine whether positive deferrals mediated the relationship with individual performance. To test for mediation, we conducted a series of three regression analyses as recommended by Baron and Kenny (1986). First, we regressed individual performance on numeric status (where solo = 1; all others = 0) multiplied by task sex-stereotypicality (where typical = 1; atypical = 0) including our control variables. The interaction of numeric status and task sex-stereotypicality was positively related to individual performance [$F(5,343) = 19.08$, $p < .01$]. Next, we regressed the mediator, positive deferrals received, on the interaction of numeric status and task sex-stereotypicality as well as our controls. The interaction of numeric status and task sex-stereotypicality was, again, significantly and positively related to positive deferrals received [$F(5,343) = 9.36$, $p < .01$]. Finally, we regressed individual performance on positive deferrals received and our controls. The number of deferrals received was positively related to performance [$F(3,343) = 36.67$, $p < .01$]. A Sobel (1982) test indicated that the number of positive deferrals received partially mediated the relationship between being a solo member working on a sex-typical task ($t = 2.37$, $p < .05$; see Fig. 3) and individual performance since the magnitude of the direct effect of the interaction of numeric status and task sex typicality on individual performance declined significantly (from 1.03 to .84) when the mediator, positive deferrals, was included in the model. Thus, hypothesis 3 was supported.

Discussion

Predicting how a person will perform in a group requires having explicit and delineated information both about the prevalence of members of their sex in that group—whether they are in the majority or minority—and the extent to which the person is a member of the sex that is viewed as typically performing and having expertise on the group's task. We concur with past research that has suggested that solo status can significantly harm perceptions by others and a person's motivation and performance, but we also depart from past research by hypothesizing and finding that solo status and a positive association between sex and a task provides a performance boost; not just for men, but for women as well. Thus, if commensurate tasks that are equivalently typical or atypical of men and women are examined, women may not always do worse as minority members of a group. Indeed, they may even perform better than their male colleagues if the task is congruent with their sex. Further, we identified positive deferrals by other group members as an underlying mediator between these two sources of stereotyping and a person's performance in a group.

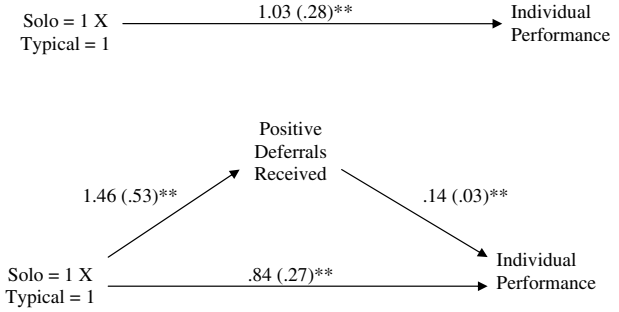

Fig. 3. The interaction of solo status and task sex-stereotypicality on individual performance as mediated by number of positive deferrals received. *Note:* Coefficients are unstandardized betas. Standard errors appear in parentheses. $^{**}p < .01$, two-tailed tests.

Positive deferrals

We first considered the joint influence of numeric status and task sex-stereotypicality on the extent to which group members positively deferred to a solo member of the other sex. Our test supported hypothesis 1A: solo men working in groups assigned to a sex-typical (math) task and solo women working in groups assigned to a sex-typical (verbal) task were significantly more likely to be deferred to by their group members than were solos of either sex working on sex-atypical tasks—verbal for men and math for women.

Hypothesis 1B expanded on 1A by predicting that not only would solo members of their sex receive significantly more or less positive deferrals from their team mates depending on the congruence of the task with their sex, but that all other combinations of numeric status and task typicality would receive intermediate levels of positive deferrals—not as many as solos working on typical tasks and not as few as solos working on atypical tasks. We found support for this hypothesis for solos working on a group task that was typical for their sex. Comparing them to all other conditions simultaneously showed that they received significantly more positive deferrals. And, in the more detailed (pairwise) comparisons they received more positive deferrals than did all other subjects except those working in homogeneous groups on a sex-atypical task. One possible explanation for this exception is that members of homogeneous groups working on atypical tasks are more likely to articulate their positive deferrals to all members of the group rather than focusing only on any single member. Taken together with the finding discussed below in which members of homogeneous groups working on sex-atypical tasks performed as effectively as solo mem-

bers working on sex-typical tasks, however, this may indicate that assigning people to same-sex groups to work on an atypical task for that sex may avoid social loafing effects since people are simultaneously comfortable with the group's composition, but aware that the task does not fit with their compositional profile and, therefore, more motivated to work hard on it and complement their colleagues for working hard. Thus, all members may have been more vigilant and felt more accountable for their work because they believed there were no "experts" in the group that could perform the task for them. Research on social loafing might further investigate how the lack of congruence between members' demographic characteristics and the task at-hand influences members' propensities to free ride in differently composed groups.

Positive deferral behaviors offer preliminary clues about how different members are treated in the group as they work together. Solo members working on sex-stereotypical tasks received more positive deferral behaviors from their colleagues than did group members in most other conditions suggesting that their colleagues encouraged their performance, and were more willing to trust and respect the validity of the answers they generated—behaviors that undoubtedly increased solo members' confidence, and thus, their ability to perform and contribute to accomplishing the group task (e.g., Steele & Aronson, 1995), which we discuss further below in our review of hypotheses 2 and 3. This supports the notion that other members reward a solo member whose sex is congruent with the task by articulating more praise and support and by deferring to their judgment.

Interestingly, we found no comparable support for the part of hypothesis 1B pertaining to solo members working in a group on a sex-atypical task. They did not receive any fewer positive deferrals compared to subjects in any other condition (except, of course, solos working on a sex-typical task), and they received even slightly *more* deferrals than members of homogeneous groups working on typical tasks. Taken together, this suggests that solo members working on typical tasks are treated with more deference than are members of groups with all other types of composition. But, those who are solos in groups working on sex-atypical tasks are treated with no less deference than are others who are not solos working on sex-typical tasks. One possible explanation for this pattern of findings and non-findings is that other group members may engage in politically correct behavior (e.g., Sommers & Norton, 2006) to ensure that they do not treat solos working in their group on sex-atypical tasks any worse than other group members, even though they do not feel that the solo member has much to contribute to the group's efforts. Members may do this by withholding deferrals generally in groups with solo members for whom the task is atypical for their sex.

Individual performance in the group

We found strong support for hypothesis 2A showing significant performance differences between solos working in groups on sex-typical tasks and solos working on atypical tasks. It is worth noting again that solo members' performance was influenced *above and beyond their actual expertise* suggesting that sex-based stereotypes, rather than a person's actual skills, contribute significantly to their behavior in groups. There were female math majors and male English majors among our subjects and yet, a solo member's actual expertise became less relevant in groups in which the solo member's sex was incongruent with the task. Thus, our study provides additional insight into the social causes of individual behavior in groups.

Solo men and women working in groups on the sex-congruent task performed significantly better than did solo men and women working in groups on the sex-incongruent task. This finding may have been accentuated by the one-time meeting of these groups dictated by our research design. More time together and greater familiarity with one another may enable groups to realize each member's true capabilities. But, while there is some evidence supporting the notion that groups learn more about people and deep level differences become more prominent than surface level differences over time (e.g., Harrison, Price, & Bell, 1998; Polzer, Milton, & Swann, 2002), there is also evidence that these stereotypes are remarkably persistent, even among intact work groups, and that they affect both perceptions and actual performance among focal minorities (Flynn et al., 2001).

In our more fine grained tests of hypothesis 2B, we also found support for our predictions. Specifically, aggregated comparisons showed that solo members working on both typical and atypical tasks performed significantly better and significantly worse, respectively, than did subjects in all other conditions. The detailed comparisons revealed three exceptions, however, in that members of balanced groups working on atypical tasks, and members of all female or all male groups working on either atypical or typical tasks performed no worse than did solos working on typical tasks. That there was no difference between members of homogeneous groups and solos working on typical tasks suggests that being the uniquely typical member provides as much of an advantage as being in a homogeneous group working on a sex-congruent task.

It also suggests that those working on atypical tasks in balanced or homogeneous groups did not exhibit social loafing or free-riding behavior. Indeed subjects who were in the majority sex working on typical or atypical tasks performed significantly worse than did both members of homogeneous groups working on typical or atypical tasks. Majority members appeared to be loafing in both the typical and atypical conditions; that

is, when their solo member was working on a sex-typical task, and unexpectedly, also when the solo member was working on the sex-atypical task. Social loafing clearly explains the first case; when the majority members were working on sex-atypical tasks it made sense for them to rely on the solo member for whom the task is sex congruent to accomplish the task. The second case, however, is harder to explain. One possibility is that majority members working on the typical task with the solo member, correspondingly, working on a task that is sex-incongruent, may have become so uncomfortable in the presence of solo atypical member that it disrupted their own performance as well. Indeed, any time a solo member was present it seemed that majority members did worse, perhaps because they paid too much attention to the minority member. This contrasts with research on stereotype lifts (Walton & Cohen, 2003) which has shown that majority members experience a performance boost as a result of downward comparisons to an inferior outgroup. A political correctness norm may have been at play here, as we suggested in our discussion of positive deferrals above (e.g., Goncalo, Chatman, & Deguid, 2008) and is worthy of further investigation to disentangle the possibility of a performance boost caused by a downward comparison from the performance decline we observed, potentially caused by unproductive distraction.

Taken together, these findings suggest that the additional scrutiny of being a solo member of one's sex, whether self-imposed or imposed by others, can be either beneficial or detrimental, depending on the typicality of the task relative to the individual's sex. If the minority member performs a task for which his or her sex is typical, such scrutiny appears to be beneficial; he or she is considered "distinctive," with all the associated positive connotations (e.g., Shih, Ambady, Richeson, Kentaro, & Gray, 2002). But, if the task is atypical for the minority member's sex, he or she is likely to perform poorly and to be considered by team mates to be, instead, "conspicuous," with all the negative connotations that term implies. Research that fails to distinguish between these sources of stereotyping—tasks and numeric status—within the immediate work group blurs relationships between demography and performance, particularly when numeric status and task sex-stereotypicality are correlated (e.g., Grier & Deshpande, 2001).

Positive deferrals and performance

In our final hypothesis we attempted to identify the mechanism underlying the strong performance exhibited by solo members working on sex-typical tasks and to better understand the relationship between positive deferrals and performance. Specifically, we found that positive deferrals partially mediated the relationship between solo status and task typicality and performance. This demonstrates the role that team members play in determining a solo member's performance in a group setting. The finding is particularly important because it highlights an *inter* personal process through which being a solo or atypical member of a group influences individuals' behavior. Past research has focused more exclusively on *intra* personal mechanisms such as stereotype threat, cognitive distraction and attention, self-consciousness, or motivation as reasons why solo or atypical status shapes individual performance. Our findings show that above and beyond these cognitive or motivational processes that occur within the individual, being a solo or atypical member can also affect performance because of processes that occur within the group in the way they interact and support, or fail to support, one another. The findings also have a practical implication; compelling a minority member to demonstrate his or her capabilities may require that other members explicitly invite his or her views more intently than if the member is not a minority.

Our study also suggests that, even when the task is stereotypically congruent with a person's sex, the solo member is subject to undue scrutiny. We already knew that this scrutiny had negative effects for solo members (e.g., Kanter, 1978), here specifically those working on sex-incongruent tasks. But, for solo members working on tasks that are stereotypic of their sex, this added attention comes in the form of positive deferrals and the solo member responds by living up to those heightened expectations. It is possible that though these solo members may gain more confidence and may be more satisfied with their group experience, they still experience more pressure from the group than do other non-solo members.

Limitations and future research

This study raises a number of implications for future research. First, our findings contradict the notion that being the numeric minority in a group is necessarily disadvantageous (Kanter, 1978). Instead, being a solo member of a group can actually enhance performance when the solo member is perceived as the unique expert with respect to groups' work. Much of the diversity literature focuses on the negative effects of separation and disparity, and suggests that groups do not make full use of variety; information resources distributed across members, including those of low (perceived) status, because of lingering status differences (Harrison & Klein, 2008). Our paper shows that task congruence can actually change the status ordering and increase the chances that groups capitalize on their inherent variety. Positive deferrals, in this light, are a way for teams to provide voice to otherwise lower status members who have critical information resources.

Our study also suggests that numeric distinctiveness, proportional representation, relational demography, and sex-stereotyped tasks are all actuarial perspectives. The three numerically focused perspectives emphasize composition in a specific context (e.g., work group, organization), while task stereotypicality focuses on longer term, macro or societal based distributions. Given the influence of both perceptions and behaviors by others as well as a person's performance in a group, tasks and demographic distributions and tasks should be examined simultaneously for their typicality across a range of social categories.

We were able to estimate the extent to which performance was influenced by numeric status and sex stereotyped tasks, above and beyond an individual's own expertise. These effects can, therefore, be largely attributed to social perceptions that may either enhance or distract a person from showing their true performance, consistent with the notion of performance boosts and identity threats (e.g., Steele & Aronson, 1995). To accomplish this we made a difficult tradeoff: on the one hand, we identified a task in which individual and group performance could be easily compared. On the other hand the task may or may not have been better accomplished by individuals versus groups. Thus, the level of interdependence we imposed may have been artificial and limited our ability to generalize these findings to, for example, groups working on less demonstrable tasks (e.g., those without a right or wrong answer such as brainstorming). Future research should attempt to identify tasks that require specific amounts of interdependence and show differential typicality for members of different social categories, but can still be assessed for individual competence.

Though we randomly assigned subjects to conditions and included a variety of controls, social categories other than sex could have been salient and influenced the relationship between numeric status, task sex-stereotypicality, and performance.[4] The principle of functional antagonism (Turner, 1985) in which the more salient any one category, the less potential for others to be salient suggests that increased salience of another category could attenuate our results. Therefore, future research should examine the relative impact of sex compared to other potential social categories. Past demography research has identified a host of such characteristics, such as tenure and function (e.g., Lawrence, 1997) and even political orientation (Spataro, 2007). The challenge would be to figure out how to combine them to understand their joint influence, as this is closer to the actual group settings that people are confronted with at work.

Our laboratory groups were small and held meetings at one point in time, placing obvious limits on the external validity of our findings as well as increasing the potential salience of surface level demographic differences such as sex more than they would have been the case in an intact team (Harrison et al., 1998). Research conducted in organizational settings may provide further insight, as in Milton and Westphal's (2005) field study of construction crews and emergency response teams. They reported that when group members experienced interpersonal congruence, that is, when they mutually reinforced one another's identities, group members were more likely to cooperate with one another. This suggests that in our study, identity confirmation through positive deferrals may be a possible explanation for why minority working on sex-congruent tasks members performed better in our study. Future researchers should continue to study these phenomena in natural settings and over time to understand how perceptions of task, job, and occupational stereotypicality originate (e.g., Harrison, Price, Gavin, & Florey, 2002).

Further, we focused on individual attitudes in this study, but there may be group level constructs that are relevant such as the group's expectations for success or sense of group efficacy on the task (Goncalo, 2004) that will also affect how inclusive the majority will be with minority members.

Numeric status and task typicality are likely linked longitudinally as well since, at some point it may become numerically rare for members of a particular social category to work on a certain task or in a certain domain. Over time, this low base rate of participation may eventually lead to the formation of stereotypes associating rarity with negative expectations (e.g., incompetence, unworthiness).

Demographic shifts in professional basketball in the US illustrate this possibility: White players dominated the National Basketball Association (NBA) until the early 1950s. Like other sports up until that time, African Americans were not allowed to play. But, African American players have now numerically dominated White players in the NBA for over 20 years. Historical analyses could track the point in time in which it became considered "typical" for African Americans to play basketball and "atypical" for Whites, and distinguish this from the numeric composition of White and African American players in the NBA. Examining a profession in which the profile of the typical incumbent has changed radically in a relatively short period would allow greater understanding of the origination and evolution of social category based stereotypes (e.g., numeric status may eventually give rise to perceptions of task typicality), and also of the correspondence between changes in numeric composition and the consequences of perceptions of task typicality. In particular, we might predict that changes in numeric composition occur far sooner than do stereotyped expectations. The challenge would be to find jobs or professions in which being in the numeric minority is not also historically atypical. Ongo-

[4] We did rule out the influence of ethnicity and year in school as alternatives by running all our models with these covariates. They were never significant, nor did they in any way change our pattern of results.

ing changes in the demographic composition of the US workforce will undoubtedly offer such opportunities, and, at the same time, increase the importance of understanding the separate and powerful effects of numeric status and task stereotypicality on perceptions, social interaction, and performance.

References

Abrams, D., Thomas, J., & Hogg, M. (1990). Numerical distinctiveness, social identity and gender salience. *British Journal of Social Psychology, 29*, 87–92.

Alexander, V. D., & Thoits, P. A. (1985). Token achievement: An examination of proportional representation and performance outcomes. *Social Forces, 64*, 332–340.

Allmendinger, J., & Hackman, J. R. (1995). The more the better? A four-nation study of the inclusion of women in symphony orchestras. *Social Forces, 74*, 423–437.

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Aronson, J., Lustina, M., Good, C., Keough, K., Steele, C., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29–46.

Babcock, L., & Laschever, S. (2003). *Women don't ask: Negotiation and the gender divide*. Princeton, NJ: Princeton University Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174–181.

Berger, J., Webster, M., Jr., Ridgeway, C., & Rosenholtz, S. (1986). Status cues, expectations, and behaviors. In E. J. Lawler (Ed.). *Advances in group processes* (Vol. 3). Greenwhich, CT: JAI Press.

Berger, J., & Zelditch, M. J. (1985). *Status, rewards, and influence*. San Francisco: Jossey Bass.

Bielby, W. T., & Baron, J. N. (1986). Men and women at work: Sex segregation and statistical discrimination. *American Journal of Sociology, 91*, 759–799.

Brewer, M. B. (1991). The social self: On being he same and different at the same time. *Personality and Social Psychology Bulletin, 17*, 475–482.

Carli, L. L., & Eagly, A. H. (1999). Gender effects on social influence and emergent leadership. In G. N. Powell (Ed.), *Handbook of gender and work*. Thousand Oaks, CA: Sage.

Cassirer, N., & Reskin, B. (2000). High hopes: Organizational position, employment experiences, and women's and men's promotion aspirations. *Work & Occupations, 27*, 438–463.

Chatman, J. A., & O'Reilly (2004). Asymmetric reactions to work group sex diversity among men and women. *Academy of Management Journal, 47*, 193–208.

Chatman, J. A., & Spataro, S. E. (2005). Using self-categorization theory to understand relational demography-based variations in people's responsiveness to organizational culture. *Academy of Management Journal, 48*, 321–331.

Cherry, F., & Deaux, K. (1978). Fear of success versus fear of gender-inappropriate behavior. *Sex Roles, 4*, 97–101.

Cohen, L., & Swim, J. K. (1995). The differential impact of gender ratios on women and men: Tokenism, self-confidence, and expectations. *Personality and Social Psychology Bulletin, 21*, 876–884.

Crocker, J., & McGraw, K. M. (1984). What's good for the goose is not good for the gander. *American Behavioral Scientist, 27*, 357–369.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Dovidio, J. F., Heltman, K., Brown, C. E., Ellyson, S. L., & Keating, C. F. (1988). Power displays between women and men in discussion of gender-linked tasks: A multichannel study. *Journal of Personality and Social Psychology, 55*, 580–587.

Driskell, J. E., Olmstead, B., & Salas, E. (1993). Task cues, dominance cues, and influence in task groups. *Journal of Applied Psychology, 78*, 51–60.

Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology, 60*, 685–710.

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

Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin, 17*, 306–315.

Elfenbein, H., & O'Reilly, C. (2007). Fitting in: The effects of relational demography and person–organization fit on group process and performance. *Group and Organization Management, 32*, 109–142.

Fairhurst, G., & Snavely, B. K. (1983). A test of the social isolation of male tokens. *Academy of Management Journal, 26*, 296–300.

Floge, L., & Merrill, D. (1986). Tokenism reconsidered: Male nurses and female physicians in a hospital setting. *Social Forces, 64*, 925–994.

Flynn, F. J., Chatman, J. A., & Spataro, S. E. (2001). Getting to know you: The influence of personality on impressions and performance of demographically different people in organizations. *Administrative Science Quarterly, 46*, 414–442.

Fottler, M. P. (1976). Attitudes of female nurses toward the male nurses: A study of occupational segregation. *Journal of Health and Social Behavior, 17*, 98–110.

Fuegen, K., & Biernat, M. (2002). Reexamining the effects of solo status for women and men. *Personality and Social Psychology Bulletin, 28*, 913–925.

Goncalo, J. (2004). Past success and convergent thinking in groups: The role of group-focused attributions. *European Journal of Social Psychology, 34*, 385–395.

Goncalo, J., Chatman, J., & Deguid, M. (2008). Political correctness and creativity in mixed and same sex groups. Working paper, Cornell University School of Industrial Relations.

Grier, S., & Deshpande, R. (2001). Social dimensions of consumer distinctiveness: The influence of social status on group identity and advertising persuasion. *Journal of Marketing Research, 38*, 216–225.

Hargreaves, D. J., Bates, H. M., & Foot, J. M. (1985). Sex-typed labeling affects task performance. *British Journal of Social Psychology, 24*, 153–155.

Harrison, D. A., & Klein, K. (2008). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review, 32*, 1199–1228.

Harrison, D. A., Price, K. H., & Bell, M. P. (1998). Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. *Academy of Management Journal, 41*, 96–107.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of diversity on group functioning. *Academy of Management Journal, 45*, 1029–1045.

Hart, J. W., Bridgett, D. J., & Karau, S. J. (2001). Coworker ability and effort as determinants of individual effort on a collective task. *Group Dynamics, 5*, 181–190.

Hecht, L. W., & Schraeder, W. B. (1986). *Technical report on test development and score interpretation for GMAT users*. Los Angeles: Graduate Management Admissions Council.

Heikes, E. J. (1991). When men are the minority: The case of men in nursing. *Sociological Quarterly, 32*, 389–401.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelman effect: Studies of group size and group performance. *Journal of Experimental Social Psychology, 10*, 371–384.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365–371.

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95*, 796–805.

Jackson, S., Joshi, A., & Erhardt, N. (2003). Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of Management, 29*, 801–830.

Kadushin, A. (1976). Men in a woman's profession. *Social Work, 21*, 440–444.

Kanter, R. M. (1978). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology, 82*, 965–990.

Karakowsky, L., & McBey, K. (2001). Do my contributions matter? The influence of imputed expertise on member involvement and self-evaluations in the work group. *Group and Organization Management, 26*, 70–92.

Karakowsky, L., & Siegel, J. P. (1999). The effects of proportional representation and gender orientation of the task on emergent leadership behavior in mixed-gender work groups. *Journal of Applied Psychology, 84*, 620–631.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681–706.

Kenny, D. A., & LaVoie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology, 48*, 339–348.

Kerr, N. L., & Brunn, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology, 44*, 78–94.

Kiesler, S. (1975). Actuarial prejudice toward women and its implications. *Journal of Applied Social Psychology, 5*, 201–216.

Kirkman, E. E., Maxwell, J. W., & Priestley, K. R. (2003). 2002 Annual survey of the mathematical sciences. *Notices of the American Mathematical Society*, 801–813.

Konrad, A. M., Winter, A., & Gutek, B. A. (1992). Diversity in work group sex composition: Implications for majority and minority workers. *Research on the Sociology of Organizations, 10*, 115–138.

Kray, L. J., Galinsky, A., & Thompson, L. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80*, 942–958.

Latane, B., Williams, K., & Harkins, S. G. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822–832.

Lau, D. C., & Murnighan, J. K. (2005). Interactions within groups and subgroups: The effects of demographic faultlines. *Academy of Management Journal, 48*, 645–659.

Lawrence, B. S. (1997). The black box of organizational demography. *Organization Science, 8*, 1–22.

Laws, J. L. (1975). The psychology of tokenism: An analysis. *Sex Roles, 1*, 51–56.

Lord, C. G., & Saenz, D. S. (1985). A meta-analysis of rate race effects in performance ratings. *Journal of Personality and Social Psychology, 49*, 918–926.

Lyness, K. S., & Thompson, D. E. (1997). Above the glass ceiling? A comparison of matched samples of female and male executives. *Journal of Applied Psychology, 82*, 359–375.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest, 6*, 31–55.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather flock together: Homophily in social networks. *Annual Review of Sociology*.

Mehra, A., Kilduff, M., & Brass, D. J. (2001). The social networks of high and low self-monitors: Implications for workplace performance. *Administrative Science Quarterly, 46*, 121–146.

Messick, D. M., & Mackie, D. M. (1989). Intergroup relations. *Annual Review of Psychology, 40*, 45–81.

Milton, L. P., & Westphal, J. D. (2005). Identity confirmation networks and cooperation in work groups. *Academy of Management Journal, 48*, 191–212.

Monahan, L., Kuhn, L., & Shaver, P. (1974). Intrapsychic versus cultural explanations for the fear of success motive. *Journal of Personality and Social Psychology, 29*, 60–64.

Neimann, Y. F., & Dovidio, J. F. (1998). Relationship of solo status, academic rank, and perceived distinctiveness to job satisfaction of racial/ethnic minorities. *Journal of Applied Psychology, 83*, 55–71.

Ott, E. M. (1989). Effects of the male–female ratio at work: Policewomen and male nurses. *Psychology of Women Quarterly, 13*, 41–58.

Pazy, A., & Oron, I. (2001). Sex proportion and performance evaluation among high-ranking military officers. *Journal of Organizational Behavior, 22*, 689–702.

Pichevin, M. F., & Hurtig, M. C. (1996). Describing men, describing women: Sex membership salience and numerical distinctiveness. *European Journal of Social Psychology, 26*, 513–522.

Polzer, J. T., Milton, L. P., & Swann, W. B. (2002). Capitalizing on diversity: Interpersonal congruence in small work groups. *Administrative Science Quarterly, 47*, 296–324.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*, 55–71.

Reskin, B. F., McBrier, D. B., & Kmec, J. A. (1999). The determinants and consequences of workplace sex and race composition. *Annual Review of Sociology, 25*, 335–352.

Sanguinetti, C., Lee, A. M., & Nelson, J. (1985). Reliability estimates and age and gender comparisons of expectations of success in sex-typed activities. *Journal of Sport Psychology, 7*, 379–388.

Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high- and low-status groups. *Personality and Social Psychology Bulletin, 28*, 694–707.

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology, 39*(1), 68–74.

Shih, M., Ambady, N., Richeson, J., Kentaro, F., & Gray, H. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 83*, 638–647.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological Methodology* (pp. 290–312). San Francisco: Jossey-Bass.

Sommers, S. R., & Norton, M. I. (2006). Lay theories about white racists: What constitutes racism (and what doesn't). *Group Processes & Intergroup Relations, 9*, 117–138.

Spangler, E., Gordon, M., & Pipkin, R. (1978). Token women: An empirical test of Kanter's hypothesis. *American Journal of Sociology, 85*, 160–170.

Spataro, S. E. (2007). Not all differences are the same: Understanding the status implications of increasing diversity in the workplace. Working paper, Cornell University.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 11). New York: Academic Press.

Thompson, M., & Sekaquaptewa, D. (2002). When being different is detrimental: Solo status and the performance of women and racial minorities. *Analyses of Social Issues and Public Policy (ASAP), 2*, 183–203.

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. *Advances in Group Processes, 2*, 77–121.

Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456–467.

Wentworth, D. K., & Anderson, L. R. (1984). Emergent leadership as a function of sex and task typicality. *Sex Roles, 11*, 513–524.

Westphal, J. D., & Milton, L. P. (2000). How experience and network ties affect the influence of demographic minorities on corporate boards. *Administrative Science Quarterly, 45*, 366–398.

Wharton, A. S., & Baron, J. N. (1987). So happy together? The impact of gender segregation on men at work. *American Sociological Review, 52*, 574–587.

Williams, C. L. (1992). The glass escalator: Hidden advantages for men in the 'female' professions. *Social Problems, 39*, 253–267.

Williams, K. Y., & O'Reilly, C. A. (1998). Forty years of diversity research: A review. In B. Staw & L. Cummings (Eds.). *Research in organizational behavior* (Vol. 20, pp. 77–140). Greenwich, CT: JAI Press.

Wood, W. (1987). Meta-analytic review of sex differences in group performance. *Psychological Bulletin, 102*, 53–71.

Yoder, J. (1991). Rethinking tokenism: Looking beyond numbers. *Gender and Society, 5*, 178–192.

Yoder, J. D. (1994). Looking beyond numbers: The effects of gender status, job prestige, and occupational gender-typing on tokenism processes. *Social Psychology Quarterly, 57*, 150–159.

Yoder, J., & Sinnett, L. M. (1985). Is it all in the numbers? A case study of tokenism. *Psychology of Women Quarterly, 9*, 413–441.

Zajonc, R. B. (1965). Social facilitation. *Science, 271*, 366–369.

Zimmer, L. (1988). Tokenism and women in the workplace: The limits of gender-neutral theory. *Social Problems, 35*, 64–77.