

DOES DIVERSITY-VALUING BEHAVIOR RESULT IN DIMINISHED PERFORMANCE RATINGS FOR NON-WHITE AND FEMALE LEADERS?

DAVID R. HEKMAN
STEFANIE K. JOHNSON
University of Colorado at Boulder

MAW-DER FOO
National University of Singapore

WEI YANG
University of Texas at Austin

We seek to help solve the puzzle of why top-level leaders are disproportionately White men. We suggest that this race- and sex-based status and power gap persists, in part, because ethnic minority and female leaders are discouraged from engaging in diversity-valuing behavior. We hypothesize, and test in both field and laboratory samples, that ethnic minority or female leaders who engage in diversity-valuing behavior are penalized with worse performance ratings, whereas White or male leaders who engage in diversity-valuing behavior are not penalized for doing so. We find that this divergent effect results from traditional negative race and sex stereotypes (i.e., lower competence judgments) placed upon diversity-valuing ethnic minority and female leaders. We discuss how our findings extend and enrich the vast literatures on the glass ceiling, tokenism, and workplace discrimination.

Women and non-Whites have made remarkable gains in the workplace in recent decades. Non-Whites and women outnumber White men in the U.S. workplace by a margin of two to one (Burns, Barton, & Kerby, 2012). However, a demographic status and power gap remains, as recent data show that only 25 Fortune 500 companies are headed by people of color and 21 by women (Catalyst, Inc., 2013; Diversity, Inc. staff, 2012). Likewise, corporate boards in the Fortune 500 are primarily composed of White men (74.4%), followed by White women (13.3%). Among ethnic minorities, 6.8% of corporate board members are African American, 3.1% are Latino, and 2.4% are Asian American (Zweigenhaft & Domhoff, 2011). The status and power gap between men and women persists despite meta-analytic evidence suggesting that women tend to be rated as better leaders than men (Paustian-Underdahl, Walker, & Woehr, 2014), and any performance evaluation gap that may exist fails to account for the highly visible status and power gap

We would like to thank Russell Cropanzano and Yuchen Zhang, as well as the entire leadership division of the Leeds School of Business. We would also like to thank Roland Smith, Michael Campbell, Shannon Bendixen, Xiuxi (Sophia) Zhao, and Angeline Lim at the Center for Creative Leadership, and, last but not least, the RSG for their tireless support of this project.

within organizations (Joshi, Son, & Roh, 2015). Despite non-Whites and women outnumbering and sometimes outperforming their White male counterparts, only rarely are they given the reigns of the most powerful organizations in society. Economists are perhaps most disturbed by this phenomenon, as orthodox economic theory would predict that it is suboptimal for society to select its top leaders from only 34% of the population (i.e., the White men; *The Economist*, 2008).¹

One way to potentially reduce this status and power gap is to place women and non-White leaders

¹ The so-called “glass ceiling” is a major reason why non-Whites and women are considered minorities, even though together they comprise a numerical majority. The term “minority” does not refer to a smaller number of people compared to the dominant group, but rather refers to a group that holds few positions of social power (Schaefer, 1996). Affirmative action programs and corporate diversity offices have been put in place with the purported goal of helping minorities break through this glass ceiling and achieve greater organizational status, power, and influence (Harrison, Kravitz, Mayer, Leslie, & Lev-Arey, 2006; Levi & Fried, 2008). However, despite the increasing emphasis on promoting diversity, ethnic minorities and women are still underrepresented at the highest organizational levels and overrepresented at the lower organizational levels (Leslie, Mayer, & Kravitz, 2014).

in high-status, high-power positions, in the hope that these leaders will empower other women and non-Whites (Ely, 1994; Ibarra, 1995; Ragins & Scandura, 1999). However, promoting demographic minorities into top leadership positions has been shown to have some unfortunate side effects. In fact, powerful minorities have been found to actively oppose the advancement of their fellow minority group members (Ellemers, Rink, Derks, & Ryan, 2012; Ely, 1994; Sheppard & Aquino, 2013), possibly because they feel threatened by fellow members of their demographic groups (Duguid, 2011; Duguid, Loyd, & Tolbert, 2012; Ely, 1994), or because, once demographic minorities break into higher-status ranks, they wish to retain their status by denying it to others (Kanter, 1977).

We suggest that another reason ethnic minority and women leaders may impede the advancement of other ethnic minorities and women is because they are penalized in the form of lower performance ratings when they engage in “diversity-valuing behavior,” or behavior that promotes demographic balance within organizations (e.g., behaviors such as hiring and promoting ethnic minorities and women). By “balance,” we mean an organizational demographic profile that comes closer to resembling the demographic makeup of the broader region or geographic area. Thus, perfect gender balance involves a 50/50 split between men and women employees, and perfect racial balance involves the same percentage of ethnic minorities working within an organization as there are in the surrounding region. When non-White and female leaders engage in diversity-valuing behavior, they are perceived as threatening the existing status and power structure (Chattopadhyay, Tluchowska, & George, 2004) and violating the expectation that minorities should play a supporting rather than a leading role in society (Sheppard & Aquino, 2013). As a result, those that perceive this to be the case take actions that preserve the established status and power hierarchy by negatively stereotyping diversity-valuing ethnic minority and women leaders as being incompetent and having low performance. Conversely, ethnic minority and women leaders avoid negative stereotypes when they engage in low levels of diversity-valuing behavior.

Our theoretical rationale linking leader demographics and diversity-valuing behavior to perceptions of competence and job performance extends and enriches the vast and expansive glass ceiling literature (e.g., Powell & Butterfield, 1994; Ragins & Scandura, 1999) by helping solve the puzzle of why the glass ceiling persists despite its societal costs. It may persist, in part, because non-White and women leaders who

engage in behaviors that increase diversity in the highest organizational ranks are systematically penalized with lower competence and performance ratings. Correspondingly, this logic may explain why there are so few leaders willing to publicly advocate for non-White or women leaders to be promoted (Ibarra, Carter, & Silva, 2010), and why ethnic minorities and women feel threatened at the prospect of hiring a fellow member of their demographic group (Duguid, 2011; Hansen, Ibarra, & Peyer, 2010). Non-White and female leaders may be highly aware of the personal danger that diversity-valuing behavior poses to their careers.

We also contribute to the literature on tokenism by providing an additional explanation for why minorities and women may impede the advancement of their fellow women (“queen bee syndrome,” Staines, Tavris, & Jayaratne, 1973: 55) and non-White (“crab mentality,” Mendoza, 2002: 57) coworkers. The tokenism literature suggests that token non-Whites and women take on the values of White men, and are placed in positions of status and power to act as gatekeepers to prevent the further dilution of those values, as well as to create the appearance of social inclusion and diversity (Kanter, 1977). Our conceptual model helps explain that token non-White or women leaders’ decision to promote White men instead of non-Whites or women could also emerge from such leaders’ awareness that diversity-valuing behavior is personally costly.

DIVERSITY-VALUING BEHAVIOR, LEADER DEMOGRAPHICS, AND PERFORMANCE RATINGS

Race and sex have been found to be two of the most important demographic markers leaders use to determine the degree to which fellow leaders are different from themselves (i.e., more important than age, education, functional background, or leadership experience; Zhu, Shen, & Hillman, 2014). In the United States, at least, most people consider White men to be members of a high-status social group and ethnic minorities and women to belong to low-status social groups (Berger, Fisek, Norman, & Zelditch, 1977). Accordingly, we follow the lead of other organizational researchers in examining both race and sex bias simultaneously (Hekman, Aquino, Owens, Mitchell, Schilpzand, & Leavitt, 2010; Zhu et al., 2014).

Although there is a great deal of evidence that a status and power gap exists for ethnic minority and women leaders at a macro level (i.e., ethnic minorities

and women are underrepresented at the highest organizational levels; Hillman, Cannella, & Harris, 2002; Hillman, Shropshire, & Cannella, 2007; Hitt & Barr, 1989; Zweigenhaft & Domhoff, 2006), there is little evidence of a main effect of race or sex on leader performance evaluations at a micro level (Eagly, Makhijani, & Klonsky, 1992; Joshi & Roh, 2009; Ng, Eby, Sorensen, & Feldman 2005; Rosette, Leonardelli, & Phillips 2008). A recent meta-analysis shows that the performance rating gap between male and female workers is miniscule compared to the status and power gap between these demographic groups (Joshi et al., 2015).

Certainly, ethnic minorities and women have made some progress in closing the status and power gap, as the percentage of Fortune 500 ethnic minority or women CEOs has doubled from 4% to 8% over the last decade (Cook & Glass, 2014; Zweigenhaft & Domhoff, 2011), and 79% of working men report having worked for a female boss at some point in their careers (Elsesser & Lever, 2011). Recent evidence suggests that ethnic minority and women leaders tend to be viewed as belonging to the “in-group” of White male top executives to the extent they are similar to the White male incumbents in other ways (i.e., similar age, leadership experience, or functional background; Zhu et al., 2014: 1), or ingratiate themselves to these incumbents (Westphal & Stern, 2006, 2007). Indeed, the relational demography literature suggests that demographically low-status group members may actually benefit from working in a group of high-status members because it increases such low-status group members’ self-esteem and social integration (Chattopadhyay et al., 2004; Chattopadhyay, George, & Ng, 2011; Van Knippenberg & Schippers, 2007).

In contrast, when low-status group members align themselves with their low-status group by engaging in diversity-valuing behavior (e.g., hiring and promoting non-White or female leaders, and respecting gender, racial, religious, and cultural differences), they could be seen as nepotistic (Wenneras & Wold, 2001) or socially competitive (Chattopadhyay et al., 2004). Individuals hold the general belief that different demographic groups are engaged in a zero-sum competition with other demographic groups, such that, if other demographic groups gain status, their own demographic groups lose status (Sidanius, Pratto, & Mitchell, 1994). Therefore, nepotism favoring non-Whites, such as affirmative action, tends to be judged even more negatively than nepotism favoring Whites, like legacy policies (Gutiérrez & Unzueta, 2013).

Diversity-valuing behavior can also be seen as socially competitive. “Social competition” involves improving the status of one’s own low-status category at the expense of high-status categories (Chattopadhyay et al., 2004). Individuals with presumed social competition motives, such as low-status individuals engaging in diversity-valuing behavior, are viewed unfavorably because they are thought to be incapable of “making it on their own” in that they must resort to getting ahead by advancing their low-status categories instead of competently performing their work (Chattopadhyay et al., 2004). However, members of low-status demographic groups who actively resist engaging in diversity-valuing behavior will be viewed as very socially uncompetitive, as such individuals avoid advancing the low-status groups to which they belong.

Further, when ethnic minority or women leaders behave in a way that highlights their low-status race or sex characteristics, their perceived low-status demographics become instantly salient, thus activating negative stereotypes associated with their low-status category (Gaertner et al., 1999; Park & Westphal, 2013). Perceived incompetence is the core of most of the stereotypes that tend to be leveled against non-Whites and women (Eagly & Steffen, 1984; Fiske, Cuddy, Glick, & Xu, 2002), particularly in work contexts (Heilman, Block, & Stathatos, 1997). It follows that, by engaging in diversity-valuing behavior, non-White and female leaders highlight their low-status demographics, which leads others to stereotype such low-status leaders as incompetent.

However, members of high-status demographic groups are also increasingly motivated through formal organizational policies to engage in diversity-valuing behavior, largely due to the growing attention to the financial (Dezsö & Ross, 2012; Herring, 2009; Lopuch & Davis, 2014; Nishii, 2013; Walker, Field, Bernerth, & Becton, 2012; Wayne & Casper, 2012) and moral (Unzueta & Knowles, 2014) benefits of organizational diversity efforts. For example, Google plans to invest \$150 million, Intel plans to invest \$300 million, and Apple plans to spend \$50 million in order to increase organizational racial and gender balance in the coming years (Kelly, 2015). Additionally, White guilt may also motivate diversity-valuing behavior by members of high-status demographic groups, as it has been shown to lead to more favorable attitudes toward one type of diversity-valuing behavior: using affirmative action policies when making hiring decisions (Swim & Miller, 1999).

Stereotyping White men who engage in diversity-valuing behavior is much more challenging than

stereotyping non-Whites and women because doing so would threaten one's beliefs about society and the status quo in which Whites and men maintain the position as highly valued members of society (Jost, Banaji, & Nosek, 2004; Magee & Galinsky, 2008; O'Brien & Major, 2005). Likewise, there are very few (if any) negative stereotypes that can be applied to White men (Fiske et al., 2002). High-status group members are afforded idiosyncrasy credit, and, therefore, given freedom to deviate from the status quo (Hackman, 1992; Hollander, 1958). Thus, it is relatively easy to stereotype demographically low-status individuals engaging in diversity-valuing behavior as incompetent (partly because this behavior highlights low-status demographics), but more difficult to stereotype their equally diversity-valuing high-status (i.e., White male) counterparts as such. As a result, diversity-valuing behavior is particularly likely to lead non-Whites and women enacting the behavior to be judged as incompetent (Fiske et al., 2002).

White and male leaders might actually benefit in terms of higher performance ratings for engaging in diversity-valuing behavior, as compared to White and male leaders who do not engage in it. Although competence judgments are the most important component of stereotypes, interpersonal warmth judgments are a close runner-up (Cuddy, Fiske, & Glick, 2008). Warmth and competence judgments account for 82% of the variance in perceptions of everyday social behaviors (Abele, Cuddy, Judd, & Yzerbyt, 2008; Cuddy et al., 2008). Specifically, warmth and competence judgments tend to be negatively correlated and "operate reciprocally, like a see-saw" for demographically low-status groups, such that enhanced warmth judgments tend to diminish competence judgments of women or non-Whites (Fiske, Xu, Cuddy, & Glick, 1999: 476). However, exuding high levels of interpersonal warmth can actually enhance competence perceptions of Whites and men (Cuddy, Fiske, & Glick, 2007). Although some observers may view diversity-valuing behavior as nepotistic and socially competitive, others may perceive it as a highly interpersonally warm behavior, as the behavior involves showing a type of compassion toward members of low-status groups by enhancing such individuals' career prospects. Taken together, if diversity-valuing behavior leads to enhanced perceptions of warmth, this judgment will lead to lower perceptions of competence and performance for non-Whites and women, but may actually enhance perceptions of competence and performance for their White or male counterparts. This logic is consistent with the

finding that demographically low-status tokens tend to be marginalized in the workplace (non-White or female CEOs remain rare), whereas demographically high-status tokens tend to reach the most powerful and highest-status organizational positions (many head nurses and school principals are White men; Barnett, Baron, & Stuart, 2000; Chattopadhyay et al., 2011; Hultin & Szulkin, 1999). Ironically, and consistent with how stereotypes operate, we argue that, regardless of whether diversity-valuing behavior is viewed unfavorably (as nepotism and social competition) or favorably (as interpersonally warm), these judgments will diminish competence and performance ratings for non-Whites and women, but not for Whites and men.

To be clear, we contend that the mechanism linking the joint effect of diversity-valuing behavior and leader demographics on performance ratings is perceived competence. Non-Whites and women tend to be stereotyped as incompetent because competence perceptions result from status perceptions, warmth perceptions, perceived social competition perceptions, and perceived nepotism perceptions (Chattopadhyay et al., 2004; Fiske et al., 2002). Thus, diversity-valuing behavior performed by low-status leaders highlights such leaders' low-status demographics; is perceived as nepotistic, socially competitive, or interpersonally warm; activates status-based competence stereotypes; and, therefore, leads to worse performance ratings. In contrast, diversity-valuing behavior performed by high-status leaders highlights such leaders' high-status demographics, and may be perceived as interpersonally warm, which might actually enhance perceptions of such leaders' competence and performance.

Hypothesis 1. Leader diversity-valuing behavior will be more negatively related to performance ratings for leaders belonging to low-status demographic groups (i.e., non-White, female) compared to leaders belonging to high-status demographic groups (i.e., White, male).

Hypothesis 2. The interactive effect of leader demographics and diversity-valuing behavior on performance ratings will be mediated by perceived competence.

METHOD

We conducted two studies to test our hypotheses. Following the full-cycle research approach (Chatman & Flynn, 2005), we first tested our conceptual model in the field, and then sought to rule out alternative

explanations and verify the direction of the causal arrow by testing our model in a highly controlled laboratory context.

STUDY 1

Sample

Our sample consisted of an entire calendar-year cohort of 362 executives working in the United States who were selected by their bosses to attend a week-long, intensive, executive development program at the Center for Creative Leadership. Missing values in some of the control variables reduced our usable sample to 350 executives. These executives were rated by their bosses and peers via a confidential online survey in which the peers and bosses reported their ratings of the executive's diversity-valuing behavior, competence, and performance two weeks prior to the start of the executive development program (100% response rate). Ten percent of the executives were non-White and 31% were female, 89% were executives (vice presidents, directors, and board-level professionals), and 11% were one rank above executives (i.e., "top managers" such as CEOs or other C-level leaders). The majority of ratees (89%) had a bachelor's degree or graduate degree. Each ratee was rated by a single boss, and an average of 3.61 peers (median = 3). Peers and bosses were demographically similar to executive ratees, as 31% of the peers were women and 14% were non-White. Thirteen percent of the executives' bosses were women and 9% were non-White.

Measures

Leader performance. We used an adaptation of Sadri, Weber, and Gentry's (2011) three-item measure of leader performance by using boss ratings on a five-point Likert scale (1 = *among the worst*, 5 = *among the best*): (1) How would you rate this person's performance in his/her present job?, (2) How effectively would this person handle being promoted in the same function or division (moving a level up)?, and (3) Where would you place this person as an executive relative to other *executives* inside and outside your organization? Coefficient α for this measure was .88.

Leader sex. We coded male leaders as 0 and females as 1. Only eight leaders were both non-White and female.

Leader race. We coded White leaders as 0 and non-White leaders as 1.

Leader diversity-valuing behavior. We used an adaptation of the Miville-Guzman's measure (Miville

et al., 1999) to capture the degree to which peers perceived the leader as engaging in diversity-valuing behaviors. Other-reports of behavior are thought to be more accurate than self-reports (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007), which is why many researchers have argued for the use of observer reports (Connelly & Ones, 2010; Oh, Wang, & Mount, 2011; Zimmerman, Triana, & Barrick, 2010), and which is why we used peer reports in our analyses. We averaged the peer ratings to create the diversity-valuing behavior score for each leader. The following three items were rated on a five-point Likert scale (1 = *not at all*, 5 = *to a very great extent*): (1) understands and respects cultural, religious, gender, and racial differences; (2) values working with a diverse group of people; and (3) is comfortable managing people from different racial or cultural backgrounds. Coefficient α for the average peer ratings of each leader of this measure was .87. The peer ratings of each leader's diversity-valuing behavior had an acceptable level of agreement, as indicated by a significant F statistic for analysis of variance (ANOVA) ($F = 2.11$, $df = 361$, $p < .01$) and r_{WG} and ICC(1, 2) statistics above acceptable cutoffs ($r_{WG} = .92$; ICC(1) = .27; ICC(2) = .52), and thus were aggregated. Aggregating data is appropriate when the F statistic for ANOVA is significant, r_{WG} is higher than .70 (James, Demaree, & Wolf, 1984), and ICC(1) is non-zero (Bliese, 2000). Certainly, ICC(2) values should be higher than .70 (Kozlowski & Klein, 2000), but a low ICC(2) value simply indicates lower power in detecting relationships involving level 2 variables (and thus a more conservative test of the hypotheses), and does not prevent aggregation (Bliese, 2000).

Perceived leader competence. We used an adaptation of Fiske and colleagues' (2002) competence scale to capture the degree to which leaders were perceived as being competent by their supervisors. Bosses rated the frequency that they believed the leader exhibited the following characteristics on a six-point Likert scale (1 = *never*, 6 = *always*): (1) effective—gets projects done well and on time, (2) impressive—one whose achievements stand out, (3) is ready for more responsibility, and (4) productive—gets a lot done. Coefficient α was .80.

Controls. Following Spector and Brannick's (2011) advice to avoid over-controlling variance, we only included those control variables expected to affect the hypothesized relationships.

- **Industry dummies and job function dummies.** Because different industries and job functions have different norms regarding the appropriate ranges of job performance ratings (Brutus, Fleenor, & London, 1998), we created dummy variables

representing the 26 industries and dummy variables representing the 20 job functions in our sample.

- *Pay.* Higher paid leaders may be viewed more favorably. The bosses were asked to report how much the leaders were paid annually.
- *Organization size.* Diversity initiatives are likely more common in large organizations, arguably making diversity-valuing less notable in these types of workplaces. Bosses were asked to report the approximate number of employees working for the organization.
- *Leader education level.* Because more educated leaders may be more savvy about avoiding being stereotyped negatively, the leaders were asked to report their highest educational degree (1 = high school; 2 = associate's; 3 = bachelor's; 4 = master's; 5 = doctorate/professional).
- *Leader organizational level.* Because it becomes more difficult to differentiate between leader quality at higher levels (Avolio, Walumbwa, & Weber, 2009), the leaders and their bosses were asked to determine each leader's organization level.² All individuals in the sample were classified by themselves and their bosses as being in the top two categories. The average leader organization level was 5.11.
- *Leader non-native English speaker.* Because English proficiency and accents have been shown to influence leader performance ratings (Neeley, 2013), leaders who reported English as their native tongue were coded 0, whereas leaders who reported being non-native English speakers were coded as 1.
- *Boss and peer familiarity with leader.* Because bosses/peers who are more familiar with ratees tend to rate these individuals as better performers (Kingstrom & Mainstone, 1985), we controlled for the variables *boss familiarity with the leader* as well as *average peer familiarity with the leader* using a single-item measure evaluated with a four-point Likert scale. This item was, "How well do

you know the ratee?" (1 = *I hardly know this person*; 2 = *I do not know this person well*; 3 = *I know this person moderately well*; 4 = *I know this person extremely well*). Because individuals tend to have a demographic similarity bias (Turban & Jones, 1988), we also controlled for the interactions of *boss race* \times *leader race*, *average peer race* \times *leader race*, *boss sex* \times *leader sex*, and *average peer sex* \times *leader sex*. We measured and coded boss/peer race (0 = White, 1 = non-White) and boss/peer sex (0 = male, 1 = female).

Results

Table 1 reports the means, standard deviations, and correlation coefficients between the criterion, predictor, and control variables.

We also ran two ANOVAs testing whether there is a race and sex difference in reported diversity-valuing behavior, and found that women are significantly more likely to be reported by their peers as engaging in diversity-valuing behavior than men (mean difference was .17, $F = 9.81$, $p < .01$), and non-Whites are significantly more likely to be reported by their peers as engaging in diversity-valuing behavior than Whites (mean difference was .36, $F = 24.26$, $p < .001$). Hierarchical moderated regression models were used to examine the hypothesized interaction effects. Following Aiken and West (1991), all variables involved in the interaction terms were mean-centered. Table 2 presents the results of the analysis for the influence of leader demographics and diversity-valuing behavior on the mediators, and Table 3 presents the results of the analysis for the influence of the main effects and proposed mediators on leader performance.

Figure 1 shows that we hypothesized two models of moderated mediation (Muller et al., 2005). Thus, in Model 1 of Table 2, all the control variables and main effects for predicting the mediator are included. In Models 2–4 (Table 2), the control variables, main effects, and hypothesized interactions predicting the mediator are included. In Model 1 of Table 3, all the control variables for predicting the criterion variable are included. In Model 2 (Table 3), all the control variables and main effects for predicting the criterion variable are included. In Models 3, 5, and 7 (Table 3), the control variables, main effects, and interactions are included for predicting the criterion variable. In Models 4, 6, and 8 (Table 3), all the control variables, main effects, interactions, and the proposed mediator of perceived leader competence are included for predicting the criterion variable.

² Organizational level was classified according to six choices (1 = hourly employee, machine operators, clerical/secretarial and support staff, technicians; 2 = first-level forepersons, crew chiefs, section supervisors; 3 = middle, office managers, professional staff, mid-level administrators; 4 = upper middle, department leaders, plant managers, senior professional staff; 5 = executives, vice presidents, directors, board-level professionals; and 6 = top managers, chief executives or operating officers, presidents).

TABLE 1
Means, Standard Deviations, and Correlations between Criterion, Predictor, and Control Variables (Study 1)

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Leader performance	4.00	.84	—														
2. Perceived competence	4.64	.49	.67	—													
3. Leader pay	\$173,824.98	88717.8	-.09	-.05	—												
4. Organizational size	4717.12	4738.15	-.02	-.02	.04	—											
5. Leader education level	3.36	.97	.01	.02	.17	.05	—										
6. Leader hierarchical level	5.11	.32	.08	.13	.06	-.25	-.01	—									
7. Leader non-native English	.04	.20	.06	.06	-.02	.13	.20	.02	—								
8. Boss familiarity with leader	3.52	.52	.23	.13	.05	-.15	-.03	-.02	.09	—							
9. Peer familiarity with leader	3.17	.35	.05	.05	.12	-.09	-.04	.12	-.05	.12	—						
10. Female boss	.13	.34	.06	.04	-.15	.07	.04	.04	-.04	-.06	-.05	—					
11. Non-White boss	.09	.28	-.02	.06	-.15	.05	.03	.04	.14	-.03	.09	.08	—				
12. Female peer percent	.31	.30	.10	.03	-.23	.00	-.03	-.03	-.06	.02	-.04	.18	.08	—			
13. Non-White peer percent	.14	.22	-.03	.05	-.14	.04	.07	-.02	.18	-.01	.01	.11	.27	.00	—		
14. Diversity-valuing behavior	4.13	.47	.15	.13	-.05	.02	.01	.16	.05	-.06	.20	.17	.13	.08	.16	—	
15. Female leader	.31	.46	.02	.07	-.16	-.05	-.11	.01	-.02	.04	.06	.15	.13	.34	.13	.16	—
16. Non-White leader	.10	.30	.01	-.01	-.11	.11	.13	.07	.46	-.06	-.04	.05	.11	-.04	.30	.18	.04

Note: N = 350 leaders; all correlations larger than .10 are significant at $p < .05$.

In Hypothesis 1, we argued that leader diversity-valuing behavior would be more negatively related to performance ratings for leaders belonging to low-status demographic groups (i.e., non-Whites, females) compared to leaders belonging to high-status demographic groups (i.e., Whites, men). Table 3 shows that the coefficient for the interaction term involving leader sex and diversity-valuing behavior was significant for leader performance ($b = -.61, p < .05$). Looking at the plots in Figure 2, diversity-valuing behavior was only negatively associated with boss ratings of leader performance for women ($b = -.40, p < .05$), not for men ($b = .21, p < .05$). Table 3 also shows that the coefficient for the interaction term involving leader race and diversity-valuing behavior was significant for leader performance ($b = -.80, p < .05$). Examining the plots in Figure 3, diversity-valuing behavior was only negatively associated with boss ratings of leader performance for non-White leaders ($b = -.59, p < .05$), not for White leaders ($b = .21, p < .05$). Therefore, Hypothesis 1 is supported.

In Hypothesis 2, we argued that the interactive effect of leader demographics and diversity-valuing behavior on performance ratings would be mediated by perceived competence. As shown in Table 2, the standardized coefficient for the interaction term involving leader sex and diversity-valuing behavior was significant for perceived leader competence ($b = -.29, p < .05$). The Aiken and West (1991) methodology demonstrated leader diversity-valuing behavior was only negatively associated with boss ratings of leader competence for women ($b = -.23, p < .05$), not for men ($b = .06, n.s.$) (see Figure 4). When perceived leader competence was entered into the model predicting leader performance rating, the strength of the coefficient for the interaction of leader sex and diversity-valuing behavior decreased from $-.63$ to $-.29$, which suggests that perceived leader competence mediated the joint influence of leader diversity-valuing behavior and sex on leader performance ratings.

To further test for mediation, we ran Sobel (MacKinnon, Warsi, & Dwyer, 1995) and bootstrap (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) analyses, and both revealed evidence of mediation. Specifically, the Sobel mediation test statistic was significant ($t = 2.64, p < .01$), indicating leader competence mediated between the interaction of leader sex \times diversity-valuing behavior and leader performance (Baron & Kenny, 1986). For bootstrapped mediation analysis (MacKinnon, Lockwood, & Williams, 2004; Schneider, Ehrhart,

TABLE 2
Influence of Leader Diversity-Valuing Behavior, Race, and Sex on Boss Perceptions of Leader Competence (Study 1)

Variables	Leader Perceived Competence			
	Model 1	Model 2	Model 3	Model 4
Constant	2.40***	2.44***	2.35***	2.39***
Industry dummies	.***	.***	.***	.***
Job function dummies	.***	.***	.***	.***
Leader pay	.00	.00	.00	.00
Organizational size (headcount)	.00	.00	.00	.00
Leader education level	.05	.05	.05	.05
Leader hierarchical level	.22*	.23*	.24*	.24*
Leader non-native English	-.15	-.14	-.20	-.19
Boss familiarity with leader	.17**	.17**	.18**	.18**
Average peer familiarity with leader	-.02	-.03	-.02	-.03
Female boss	.03	.01	.04	.02
Non-White boss	.02	.05	.01	.04
Female boss × Female leader	.33*	.36*	.34*	.38*
Non-White boss × Non-White leader	.55*	.51*	.64*	.60*
Female peer percent	.00	.00	-.02	-.02
Non-White peer percent	.19	.21	.19	.20
Female peer percent × Female leader	.44*	.46*	.48*	.50*
Non-White peer percent × Non-White leader	-1.30**	-1.26**	-1.26**	-1.22**
Diversity-valuing behavior	.08	.07	.07	.06
Female leader	.09	.09	.09	.09
Non-White leader	.16	.13	.27*	.24*
Diversity-valuing behavior × Female leader		-.29*		-.29*
Diversity-valuing behavior × Non-White leader			-.49*	-.49*
Adjusted R^2	.11	.12	.12	.13
R^2	.27	.28	.29	.30
Change in R^2		.01*	.02*	.03*

Note: $N = 350$ leaders.

* $p < .05$

** $p < .01$

*** $p < .001$, unstandardized coefficients are presented

Mayer, Saltz, & Niles-Jolly, 2005), we created 1,000 bootstrap samples and thus 1,000 estimates of the mediated effect. This analysis tested for the joint effect of leader sex and diversity-valuing behavior on leader performance rating through the mediator of leader competence. The overall indirect effect was $-.33$ (the 95% confidence interval for the mediated effect ranged from $-.66$ to $-.05$ and did not straddle zero), indicating that the moderated mediating effect was significant ($p < .05$). Thus, Hypothesis 2 is supported for sex.

For the race effect, the coefficient for the interaction term involving leader race and diversity-valuing behavior was significant for leader competence ($b = -.49$, $p < .05$; Table 2). The Aiken and West (1991) methodology demonstrated leader diversity-valuing behavior was only negatively associated with boss ratings of leader competence for non-White leaders ($b = -.43$, $p < .05$), not for White leaders ($b = .06$, n.s.) (see Figure 5). Moreover, when perceived leader

competence was entered into the model predicting leader performance rating, the strength of the coefficient for the interaction of leader race and diversity-valuing behavior decreased from $-.80$ to $-.26$, which suggests that perceived leader competence mediated the joint influence of leader diversity-valuing behavior and race on leader performance ratings (Baron & Kenny, 1986).

To further test for mediation of this race effect, we ran Sobel (MacKinnon et al., 1995) and bootstrap (MacKinnon et al., 2002) analyses and both revealed evidence of mediation. Specifically, the Sobel mediation test statistic was significant ($t = 2.31$, $p < .05$), and the overall indirect effect across 1,000 bootstrapped samples was $-.54$ (the 95% confidence interval for the mediated effect ranged from $-.98$ to $-.10$ and did not straddle zero), indicating that the moderated mediating effect was significant ($p < .05$). Therefore, Hypothesis 2 is supported for both sex and race.

TABLE 3
Influence of Leader Diversity-Valuing Behavior, Race, and Sex and Boss Judgments of Leader Competence on Boss Ratings of Leader Performance (Study 1)

Variables	Leader Performance							
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Constant	.79 ***	.19 ***	.29 ***	-2.41** ***	.14 ***	-2.49** ***	.23 ***	-2.41** ***
Industry dummies	***	***	***	***	***	***	***	***
Job function dummies	***	***	***	***	***	***	***	***
Leader pay	.00	.00	.00	.00	.00	.00	.00	.00
Organizational size (headcount)	.00	.00	.00	.00	.00	.00	.00	.00
Leader education level	.03	.02	.02	-.04	.03	-.03	.02	-.03
Leader hierarchical level	.26	.20	.21	-.05	.22	-.05	.23	-.04
Leader non-native English	.09	-.13	-.10	.06	-.19	.03	-.17	.04
Boss familiarity with leader	.40***	.43***	.43***	.24***	.44***	.24***	.44***	.24***
Average peer familiarity with leader	.01	-.04	-.05	-.02	-.04	-.01	-.05	-.02
Female boss	.09	.07	.03	.02	.09	.05	.05	.03
Non-White boss	-.20	-.20	-.14	-.20	-.22	-.23	-.16	-.20
Female boss × Female leader	.06	.07	.14	-.28	.08	-.30	.15	-.26
Non-White boss × Non-White leader	1.23	1.21	1.11	.55	1.36	.64	1.26	.60
Female peer percent	.12	.15	.16	.17	.15	.17	.14	.16
Non-White peer percent	.06	-.01	.01	-.22	-.02	-.24	.00	-.22
Female peer percent × Female leader	.66*	.63	.68*	.21	.75	.21	.79*	.24
Non-White peer percent × Non-White leader	-1.64**	-2.02**	-1.94**	-.54	-1.96**	-.55	-1.88**	-.53
Diversity-valuing behavior		.24*	.22*	.14	.23*	.15	.21*	.14
Female leader		-.10	-.10	-.19*	-.10	-.19*	-.10	-.19
Non-White leader		.29	.23	.08	.46*	.16	.40	.14
Diversity-valuing behavior × Female leader			-.63*	-.29*			-.61*	-.29*
Diversity-valuing behavior × Non-White leader					-.80*	-.26*	-.80*	-.26*
Perceived leader competence				1.11*		1.12*		1.10*
Adjusted R ²	.10	.12	.14	.49	.13	.49	.15	.49
R ²	.26	.28	.30	.58	.29	.58	.31	.59
Change in R ²		.02*	.02**	.28***	.01*	.29***	.03**	.27***

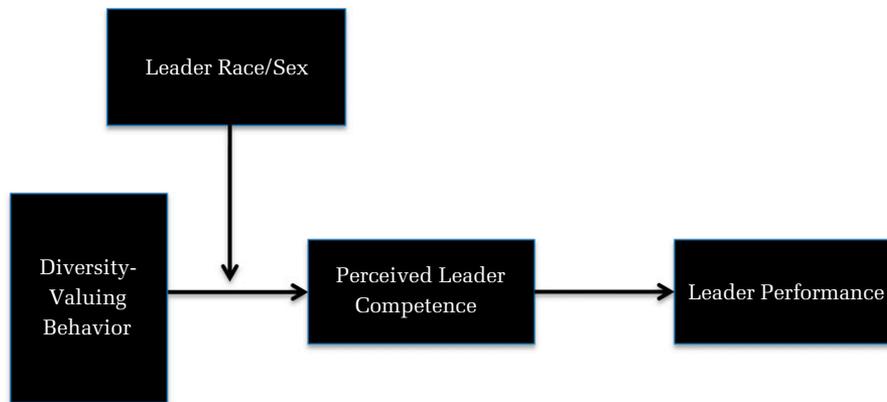
Note: N = 350 leaders.

* p < .05

** p < .01

*** p < .001, unstandardized coefficients are presented

FIGURE 1
Conceptual Model of How Leader Diversity-Valuing Behavior and Demographics Jointly Influence Ratings of Leader Competence and Performance



Discussion

Study 1 explored whether leaders are judged differently for diversity-valuing behavior depending on their demographic characteristics. Diversity-valuing behavior was only negatively related to evaluations of leaders who were non-White or female. This finding suggests that minorities and women might be able to advance their own careers by engaging in lower levels of diversity-valuing behavior. We argue that diversity-valuing women and non-Whites are rated lower than their non-diversity-valuing counterparts because diversity-valuing behavior activates subtle and unconscious stereotypes about women and non-Whites as being less competent. Therefore, biases against diversity-

valuing minority employees creep into performance evaluations.

Somewhat surprisingly to us, non-White executives were rated more favorably than their White counterparts in Study 1 (i.e., a positive main effect of being non-White on competence ratings). This may have resulted because tokens tend to need to actually perform better than non-tokens in order to advance into the highest organizational levels (such as the context for Study 1; Lyness & Thompson, 2000; Ng et al., 2005). Members of demographic groups that comprise less than 15% of an organization take on a token status (Ely, 1994), which means that non-Whites were tokens in the Study 1 sample (they comprised only 10% of the sample) but women were

FIGURE 2
Interactive Effect of Leader Diversity-Valuing Behavior and Sex on Ratings of Leader Performance (Study 1)

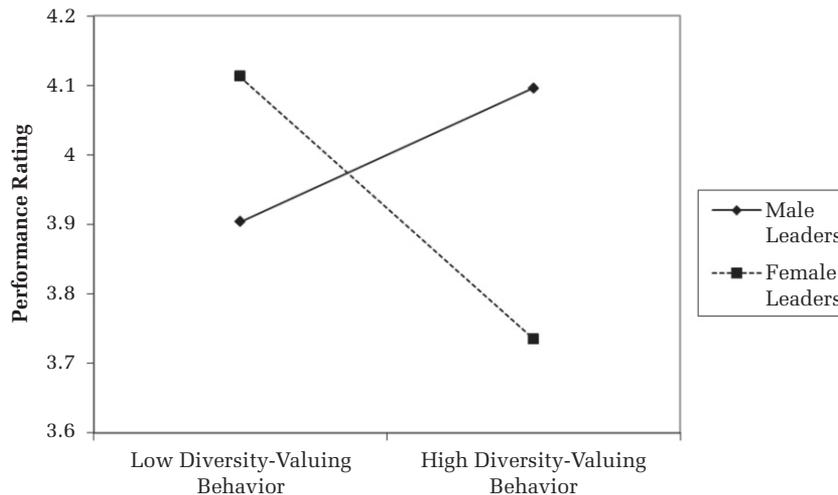
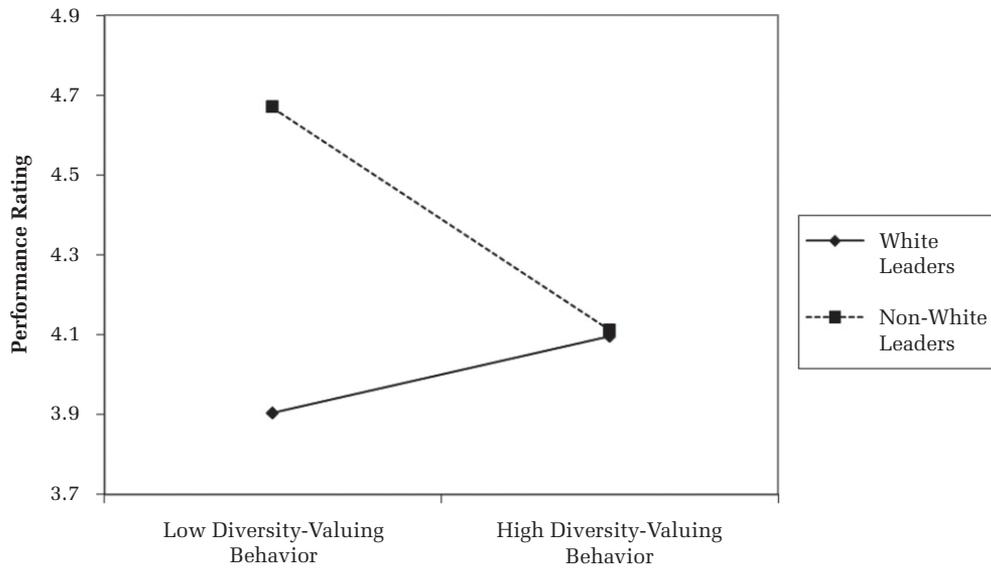


FIGURE 3
Interactive Effect of Leader Diversity-Valuing Behavior and Race on Ratings of Leader Performance (Study 1)



not tokens (they comprised 31% of the sample). As such, it is not surprising that the non-Whites, but not the women, in Study 1 received higher competence ratings, because the non-White executives were tokens and thus likely had to be more competent than their White counterparts in order to reach that high organizational level. Theoretically, we are encouraged that, despite the higher overall competence ratings for non-White executives in Study 1, we still observed our hypothesized results, which increases our confidence in our conceptual model.

Certainly, our first study has several methodological strengths, including the realism of a field study, and the leaders having been drawn from a range of industries, organizations, and functions, as well as multiple respondents from two sources (peers and supervisors). However, our findings from this study are also subject to at least two major alternative explanations. The first is that our results might be explained simply by White men actively working together in the form of an “old boys’ network” to limit the mobility of women and non-Whites who threaten

FIGURE 4
Interactive Effect of Leader Diversity-Valuing Behavior and Sex on Perceived Leader Competence (Study 1)

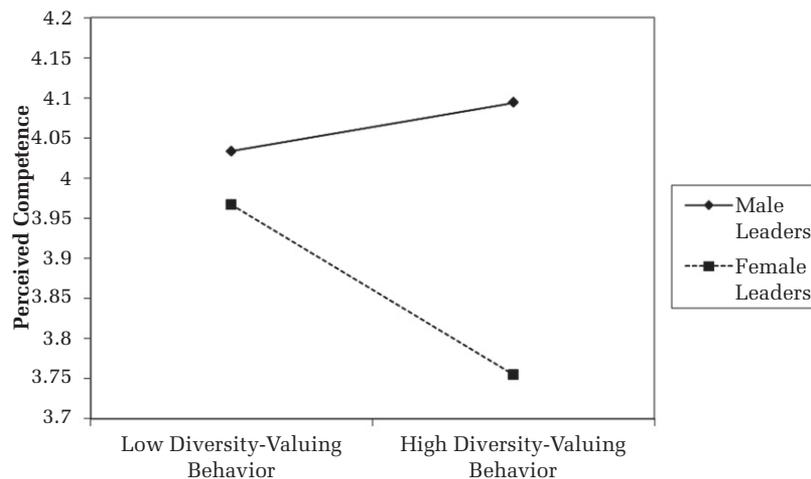
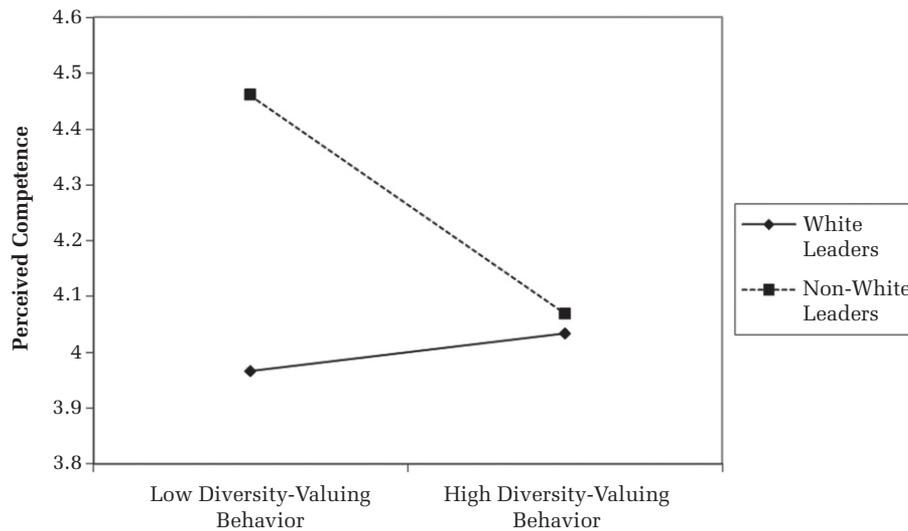


FIGURE 5
Interactive Effect of Leader Diversity-Valuing Behavior and Race on Ratings of Perceived Leader Competence (Study 1)



the status quo (Brass, 1985; Watkins et al., 2006). Organizations with powerful White men scoring highly in neosexism (Swim, Aikin, Hall, & Hunter, 1995) and modern racism (McConahay, 1986; Sears, 1988) will likely resist organizational efforts to advance race and gender balance and thus oppose the advancement of women and non-Whites who engage in diversity-valuing behavior. If these White men share and talk about their modern racist and sexist values, they may take collective actions that block diversity-valuing non-Whites and women from advancing. Further, if the values of the “old boys’ network” influence organizational values, support for non-Whites who do not engage in diversity-valuing behavior could be the result of organization-wide values or norms aimed at maintaining the status quo. Study 2 helps eliminate this alternative explanation because its participants were randomly sampled from a large online database, they were employed by a large variety of organizations spread across North America, they had nothing to gain in terms of increased personal power or prestige by behaving in a socially undesirable manner (i.e., penalizing diversity-valuing non-Whites and women with lower competence and performance ratings), and had no way to communicate or coordinate with other participants (and thus were unable to form an “old boys’ network”).

Reverse causality could also be an explanation for our Study 1 results, as incompetent non-Whites and women may be motivated to engage in a higher level

of diversity-valuing behavior in order to “stack the deck” with fellow demographically low-status individuals who may take a more lenient view of their own incompetence. Study 2 helps eliminate this alternative explanation because diversity-valuing behavior and leader demographics were manipulated, and thus we can strongly infer the direction of our conceptual model’s causal arrow. Specifically, we conducted a second study in a highly controlled laboratory context where we manipulated our predictor variables of diversity-valuing behavior as well as leader demographics.

STUDY 2

The goal of this study was to examine whether one of the most highly consequential diversity-valuing behaviors—advocating to hire a non-White or female manager—caused diversity-valuing women and non-White leaders to be negatively stereotyped and receive lower performance evaluations. To that end, we designed an experiment wherein participants were asked to read a packet of materials describing a hiring manager’s reason(s) for choosing one of two different job candidates for a vacant senior vice president position. We manipulated the hiring manager’s diversity-valuing behavior as well as the hiring manager’s demographics.

Both job candidates were equally qualified but differed based on their demographics. One of the job candidates (Candidate 1) belonged to a low-status

demographic group (i.e., female, non-White) and the other candidate (Candidate 2) was a White man. We varied the low-status demographic candidate to cover a variety of demographic groups (White female, Asian American male and female, and African American male and female). To test our hypotheses, we manipulated the demographics of the hiring manager to cover both high- and low-status demographic groups (White male and female, Asian American male and female, African American male and female) and manipulated whether or not the hiring manager engaged in diversity-valuing behavior.

Sample

Three hundred and seven adults employed in the United States participated in this study. Participants were asked to evaluate a hiring decision for a vacant senior vice president position and evaluate the competence and job performance for the hiring manager making the decision. In the final sample, 41.0% of the participants were women and 30.6% of our participants were non-White.

Design

To determine whether demographically low-status leaders receive worse competence and performance ratings when engaging in diversity-valuing behavior, we used a $2 \times 2 \times 2$ between-subjects design in which we manipulated leader sex (0 = male manager, 1 = female manager), race (0 = White, 1 = non-White), and diversity-valuing behavior (0 = no diversity-valuing behavior, 1 = diversity-valuing behavior). Participants read a scenario explaining that a four-person top management team had a split decision on which candidate to hire for an executive-level job. Participants were shown photos of the two candidate finalists. The first candidate was always a woman or ethnic minority and the second candidate was always a White man. As noted above, Candidate 1 was alternated between an Asian American woman, an Asian American man, an African American woman, an African American man, or a White woman.³

We manipulated hiring manager sex and race using high-resolution color photos of the hiring

managers. Participants were told this was the senior-ranking hiring manager who makes the final decision on who to hire. The photos of the senior hiring manager were altered to be either a White man or one of the demographically low-status groups (Asian American woman, Asian American man, African American woman, African American man, or a White woman). The hiring manager photo always either matched the demographics of Candidate 1 (e.g., the hiring manager photo and the Candidate 1 photos were both Asian American men, or Asian American women, or African American women, etc.) or matched the demographics of Candidate 2 (i.e., both the hiring manager and Candidate 2 photos were of White men).⁴

We operationalized diversity-valuing behavior as the hiring manager choosing a White male candidate versus choosing a minority or woman candidate and publicly stating that s/he valued diversity (0 = not advocating for diversity and choosing the White male candidate; 1 = publicly advocating for diversity and choosing the demographically low-status candidate).

Measures

Leader performance. Our measure of leader performance was an average score of the same three items used to measure leader performance in Study 1. Participants were asked to rate on a five-point Likert scale (1 = *among the worst*, 5 = *among the best*): (1) How would you rate this person's performance in his/her present job? (2) Where would you place this person as a leader relative to other leaders? (3) Where would you place this person as an executive relative to other executives inside and outside your organization? Coefficient α for this measure was .82.

³ We did not include Hispanic male and female photographs because we found, in pilot studies, that participants could not consistently distinguish between White and Hispanic faces.

⁴ In pilot studies, we found that demographically low-status hiring managers were penalized with lower competence and performance ratings when they advocated for demographically low-status candidates, regardless of whether there was a direct match between the low-status demographics of the hiring manager and the low-status candidate. For example, White women received an equivalent performance rating penalty whether they advocated for a White woman candidate, an African American man, African American woman, Asian American man, or Asian American woman. Thus, for sake of methodological parsimony, we designed the study so that the hiring manager demographics always matched the demographics of one of the two candidates.

Perceived competence. We used Cuddy et al.'s (2007) measure of perceived competence of the managers. Participants were asked to rate on a five-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*) the degree to which participants perceived the hiring manager as competent, confident, capable, efficient, intelligent, and skillful. Coefficient α for this measure was .91. This measure of competence was correlated at .86, $p < .001$ with the measure used in Study 1.

Leader demographic low-status manipulation. We had two between-subjects hiring manager demographic manipulations 2 (0 = male manager, 1 = female manager) \times 2 (0 = White manager, 1 = non-White manager). Manager sex and race were manipulated by using high-resolution color photos of the managers.

Leader diversity-valuing behavior manipulation. In addition, we used a between-subjects manipulation to create two diversity-valuing conditions: one in which the hiring manager advocated for hiring the White male candidate based on test scores alone, and one in which the hiring manager advocated for hiring a non-White or female candidate based on test scores as well as a desire to increase organizational racial and gender balance. The hiring manager scripts for the control and diversity-valuing conditions are shown below, and, importantly, we used our Study 1 definition of "diversity-valuing behavior" (e.g., promoting racial and gender balance) to manipulate diversity-valuing behavior in this experiment:

Control script. "Enough discussion. My reasons for choosing [the White male candidate] were the most sound, so I am going to make the final call. Candidate 2 had the highest scores and so we are going with Candidate 2."

High diversity-valuing behavior script. "Enough discussion. My reasons for choosing [the woman or ethnic minority candidate] were the most sound, so I am going to make the final call. Candidate 1 had the highest scores *and increases the racial and gender balance of our leadership team* so we are going with Candidate 1." (bracketed text and emphasis added)

To further ensure that our diversity-valuing behavior manipulation mapped onto our Study 1 diversity-valuing measure, we included the Study 1 diversity-valuing measure in Study 2. There was a significant difference between the control and diversity-valuing conditions on the Study 1 measure of diversity-valuing behavior ($F = 29.66$, $p < .001$, $M_{\text{control}} = 3.33$, $SD_{\text{control}} = .83$, $M_{\text{diversity-valuing}} = 3.95$, $SD_{\text{diversity-valuing}} = .81$).

Controls. We controlled for participant sex (0 = male, 1 = female) and race (0 = White, 1 = non-White), given the focus of the study on sex and race. Because being a non-native English-speaker could lead to participant misunderstanding the text, we controlled for participant years speaking English. We also controlled for participant age, to account for any age differences in perceptions of the task.

Results

A univariate ANOVA was used to test the interactive effects of manipulated leader diversity-valuing behavior with leader sex and leader race. Table 4 presents the means, standard deviations, and correlations between the study variables, and Table 5 presents the ANOVA results we used to test our hypotheses. Table 6 has the means and standard deviations for each cell and simple t -tests demonstrating which cell means differ significantly from one another.

TABLE 4
Means, Standard Deviations, and Correlations between Criterion, Predictor, and Control Variables (Study 2)

Variables	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1. Leader performance	3.78	.84	—							
2. Leader perceived competence	3.98	.71	.83	—						
3. Manipulated leader diversity-valuing	.51	.50	-.10	-.11	—					
4. Manipulated leader female	.67	.47	-.02	-.01	.03	—				
5. Manipulated leader non-White	.70	.46	.02	.03	-.01	-.46	—			
6. Non-White participant	.31	.46	.00	-.02	-.01	-.03	.02	—		
7. Female participant	1.41	.49	.02	.04	-.04	-.02	-.04	.02	—	
8. Participant years speaking English	33.12	9.47	.06	.05	-.06	-.07	.03	-.24	.12	—
9. Participant age	33.49	9.38	.03	.02	-.05	-.07	.05	-.23	.12	.61

Note: All correlations larger than .11 are significant at $p < .05$.

TABLE 5
Interactive Effects of Manipulated Leader Sex, Race, and
Diversity-Valuing Behavior on Perceptions of Leader
Performance and Competence (Study 2)

Variables	Leader Performance		Leader Perceived Competence	
	<i>F</i>	η^2_p	<i>F</i>	η^2_p
Non-White participant	.02	.00	437.10***	.60
Female participant	.07	.00	.05	.00
Participant years speaking English	12.61***	.04	.47	.00
Participant age	11.78***	.04	12.52***	.04
Manipulated leader female	.48	.00	12.05***	.04
Manipulated leader non-White	1.66	.01	1.609	.01
Manipulated diversity-valuing behavior	5.37*	.02	.95	.00
Manipulated diversity-valuing behavior × leader female	4.08*	.01	6.48*	.02
Manipulated diversity-valuing behavior × leader non-White	4.99*	.02	5.58*	.02

Notes: *N* = 307. Sex, race, and diversity-valuing behavior were entered into the model such that a positive value indicates a positive effect for women, non-Whites, and diversity-valuing behavior.

* *p* < .05

** *p* < .01

*** *p* < .001

In Hypothesis 1, we postulated that leader diversity-valuing behavior would be more negatively related to performance ratings for leaders belonging to low-status demographic groups (i.e., non-Whites, females),

compared to leaders belonging to high-status demographic groups (i.e., Whites, men). As shown in Table 5, the coefficient for the interaction term involving manipulated leader sex and diversity-valuing behavior on leader performance ($F = 4.08, p < .05$) and manipulated leader race and diversity-valuing behavior on performance ($F = 4.99, p < 0.05$) were both significant. As shown in Table 6 and Figure 6, the mean performance rating for male managers who valued diversity was no different than for those male managers who did not ($M = 3.75, SD = 1.00; M = 3.76, SD = .85$). Yet, when female managers valued diversity, they were rated significantly lower than when they did not ($M = 3.65, SD = .69; M = 3.96, SD = .78$), consistent with our hypothesis. The effects were similar for leader race. Among the White managers, diversity-valuing behavior had no effect on performance ratings, whereas, among the non-White managers, diversity-valuing behavior had a negative effect on performance ratings. Table 6, as well as Figure 7, presents the estimated marginal means of each cell for performance ratings. As can be seen in Figure 7, the mean performance rating for White managers who valued diversity was no different than those White managers who did not ($M = 3.73, SD = .81; M = 3.77, SD = .81$). However, consistent with our hypothesis, when non-White managers valued diversity, they were rated significantly lower than when they did not ($M = 3.61, SD = .97; M = 4.10, SD = .80$).

Hypothesis 2 was that perceived competence would mediate the demographic group by diversity-valuing behavior interaction on performance. Table 5 shows that the interaction between manipulated

TABLE 6
Cell Means and Standard Deviations (Study 2)

Leader	Manipulated Diversity-Valuing Behavior	Leader Performance		Leader Perceived Competence	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Male leader	No	3.76 ^a	.85	3.94 ^a	.76
Male leader	Yes	3.75 ^a	1.00	3.93 ^a	.82
Female leader	No	3.96 ^b	.78	4.17 ^b	.61
Female leader	Yes	3.65 ^a	.69	3.86 ^c	.59
White leader	No	3.77 ^a	.81	3.99 ^a	.71
White leader	Yes	3.73 ^a	.81	3.93 ^a	.66
Non-White leader	No	4.10 ^b	.80	4.23 ^b	.64
Non-White leader	Yes	3.61 ^a	.97	3.80 ^a	.84

Note: *N* = 307.

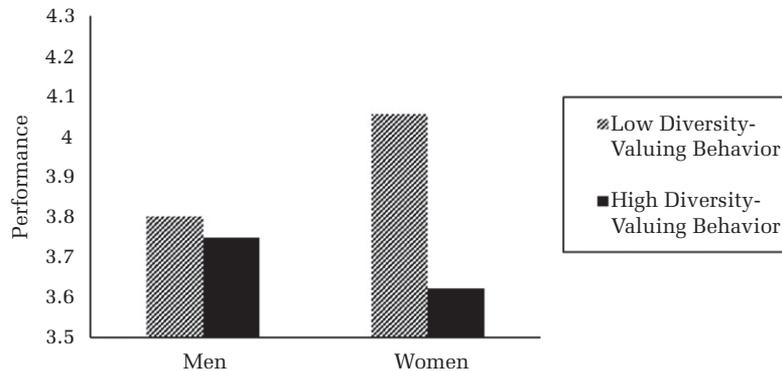
^{a, b, c} Cell means with the same superscripted letter are not significantly different from one another based on simple independent samples *t*-test (equal variances not assumed) with no control variables.

* *p* < .05

** *p* < .01

*** *p* < .001

FIGURE 6
Interactive Effect of Leader Diversity-Valuing Behavior and Sex on Ratings of Leader Performance (Study 2)



leader sex and diversity-valuing behavior on perceived competence ($F = 6.48, p < 0.05$) and manipulated leader race and diversity-valuing behavior on perceived competence ($F = 5.58, p < 0.05$) significantly predicted ratings of perceived competence. Table 6 and Figure 8 show that the mean competence rating for male managers who valued diversity was no different than that for male managers who did not ($M = 3.93, SD = .82; M = 3.94, SD = .76$). However, female managers who valued diversity were rated as significantly less competent than female managers who did not ($M = 3.86, SD = .59; M = 4.17, SD = .61$). Table 6 and Figure 9 present the estimated marginal means of each cell for competence ratings. The mean competence rating for White managers who valued diversity was no different than that for those White managers who did not ($M = 3.93, SD = .66; M = 3.99, SD = .71$). However, consistent with our hypothesis, when non-White managers valued diversity, they were rated

significantly lower on competence than when they did not ($M = 3.80, SD = .84; M = 4.23, SD = .64$).

Finally, we tested the interactive effects of demographic group and diversity-valuing behavior on performance through perceived competence using a bootstrapping methodology (MacKinnon et al., 2002) to test our moderated mediated effects with the PROCESS macro. Specifically, the bootstrapped overall indirect effect of manipulated manager sex on performance across 1,000 bootstrapped samples was $-.36$ (the 95% confidence interval for the mediated effect ranged from $-.07$ to $-.67$ and did not straddle zero), indicating that the moderated mediating effect was significant ($p < .05$). Likewise, the bootstrapped overall indirect effect of manipulated manager race on performance across 1,000 bootstrapped samples was $-.37$ (the 95% confidence interval for the mediated effect ranged from $-.01$ to $-.75$ and did not straddle zero), indicating that the moderated mediating effect was significant ($p < .05$).

FIGURE 7
Interactive Effect of Leader Diversity-Valuing Behavior and Race on Ratings of Leader Performance (Study 2)

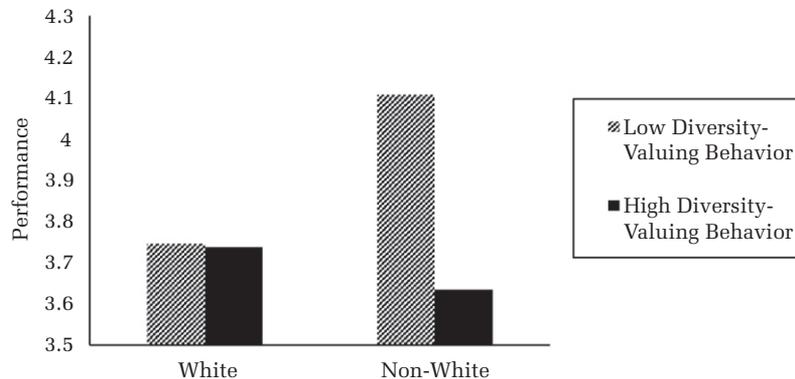
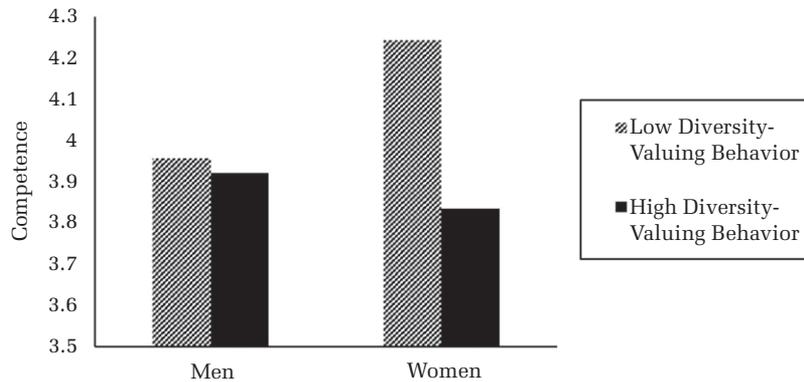


FIGURE 8
Interactive Effect of Leader Diversity-Valuing Behavior and Sex on Perceived Leader Competence (Study 2)



Thus, Hypothesis 2 is supported for both manipulated sex and race.

negative evaluations (Chattopadhyay et al., 2004; Gaertner et al., 1999).

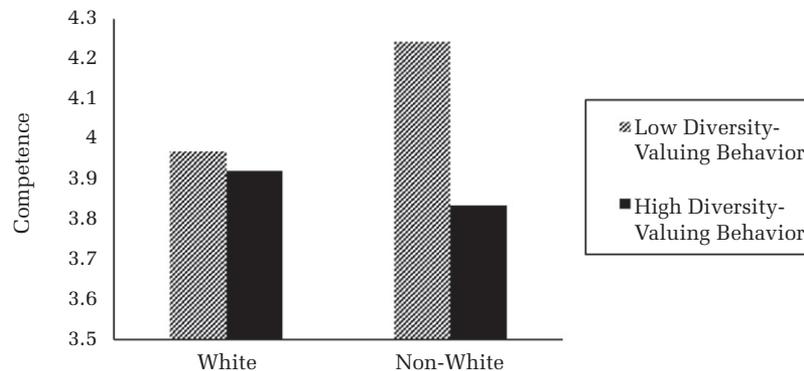
OVERALL DISCUSSION

We set out to determine whether penalties against non-White and women leaders for engaging in diversity-valuing behavior may serve to reinforce the glass ceiling. Across two samples (field and laboratory), we found clear and consistent evidence of our conceptual model, suggesting that ethnic minorities and women who engage in diversity-valuing behavior tend to be negatively stereotyped, and, thus, receive lower competence and performance ratings. Racial minorities and women who engage in diversity-valuing behavior may activate negative stereotypes associated with their group and convey a social-competition motive, resulting in more

Theoretical Implications

The two studies reported herein extend and enrich the expansive glass ceiling literature. While there are many studies documenting the glass ceiling's existence, the majority of these impressive studies are descriptive, in the sense that they have shown that the glass ceiling is a real phenomenon and pervasive across a wide range of organizations (Eagly & Karau, 2002; Hoobler, Wayne, & Lemmon, 2009; Lyness & Thompson, 1997; O'Brien, Biga, Kessler, & Allen, 2010). Likewise, another subset of the glass ceiling literature has sought to identify career strategies that individual women and non-Whites have used to break through it (Hoobler, Hu, & Wilson, 2010;

FIGURE 9
Interactive Effect of Leader Diversity-Valuing Behavior and Race on Ratings of Perceived Leader Competence (Study 2)



Lyness & Thompson, 2000; Ragins, Townsend, & Mattis, 1998). Less work has examined how the glass ceiling is maintained despite the fact that so many organizational and societal programs have been put in place to remove it (i.e., diversity initiatives, mentoring programs; Freeman, Aquino, & McFerran, 2009; Parker, Baltes, & Christiansen, 1997; Ryan & Haslam, 2005; Zoogah, 2010). In this study, we took a different perspective and tried to connect the dots between diversity-valuing behavior, stereotypes, and the glass ceiling's existence. Engaging in behaviors that increase organizational diversity hurts non-White and female leaders, in the sense that minority leaders who engage in diversity-valuing behavior fall victim to negative stereotypes, even though this behavior does not appear to harm White or male leaders.

Our findings also inform this literature by highlighting how leader demographics interact with diversity-valuing behavior to predict leader performance ratings. Observers (e.g., bosses providing performance reviews) have been found to rely most heavily on stereotypical race- and sex-based judgments when defining employee performance is highly ambiguous (Bertrand & Mullainathan, 2004; Goldin & Rouse, 2000) or when a particular work role tends to be dominated by White men (Duguid et al., 2012; Joshi & Roh, 2009). The job of a business executive is characterized by both of these attributes—that is, defining and objectively measuring executive performance is extremely difficult (Tosi, Werner, Katz, & Gomez-Mejia, 2000), and approximately 85% of executives are White men (Catalyst, Inc., 2012). Although we found only slight main effects of leader race and sex on performance evaluations in Study 1, we also found that minority and female leaders are denigrated when they engage in behaviors that may serve to activate stereotypes associated with their low status.

By identifying bosses' perceptions of leader competence as a mechanism linking the interactive effect of leader demographics and diversity-valuing behavior on performance ratings, we also inform the literature on race and sex biases. Uncovering this mechanism is a novel contribution, as previous research has focused on examining how lower performance ratings might result from an ethnic minority or female leader's perceived incongruity with the leadership role (Eagly & Karau, 2002; Johnson, Murphy, Zewdie, & Reichard, 2008; Rosette et al., 2008), or a leader's perceived dissimilarity from the rater (Joshi, 2014; Sackett & DuBois, 1991). We extend this line of theorizing by incorporating research on stereotype content (Cuddy et al., 2008) and show that engaging in diversity-valuing behavior may

undermine ethnic minority and female leaders' performance ratings because it makes them appear less competent.

Another contribution we make to the organizational literature is our conceptualization of the diversity-valuing behavior construct. Although past theorizing has noted the risks associated with advocating for one's own group (e.g., Kanter, 1977), the construct of diversity-valuing behavior has not been formally introduced or tested. Behavior that increases organizational racial and gender balance is likely to become increasingly important as organizations seek to reflect the broader societies in which they operate. However, as attribution theory would suggest, engaging in diversity-valuing behavior may result in negative attributions toward the leader, depending on his or her inferred motives for the diversity-valuing behavior (Martinko, Harvey, & Dasborough, 2011). Consistent with theorizing on social mobility and social competition, if observers infer a social competition motive for engaging in diversity-valuing behavior, they will make more negative attributions for diversity-valuing behavior (Chattopadhyay et al., 2004; Dorgan & Grieco, 1993). Our findings also add to Heilman and colleagues' (1997) work on the stigma of incompetence associated with affirmative action. Just as they found that being the recipient of affirmative action results in lower perceptions of competence, we found that low-status managers' efforts to promote others in the name of diversity resulted in lower perceptions of their competence and performance.

We contribute to the understanding of in-group and out-group favoritism (Ellemers, Heuvel, Gilder, Maass, & Bonvini, 2004) that shows female and non-White leaders also devalue low-status leaders during evaluation and promotion (Lewis & Sherman, 2003). Consistent with the relational demography literature, we find that low-status group members are motivated to work with high-status group members rather than their in-group (Chattopadhyay et al., 2004; Van Knippenberg & Schippers, 2007). As such, members of low-status groups (i.e., women and ethnic minorities) advance the social standing of members of high-status groups (i.e., White men; Chattopadhyay et al., 2004; Derks, Ellemers, van Laar, & de Groot, 2011) because advancing high-status group members is thought to provide low-status group members with a psychic reward in terms of a sense of belonging among the White male demographic elite (Bettencourt, Charlton, Dorr, & Hume, 2001; Chattopadhyay et al., 2004; Joshi, 2014). We add to that literature by suggesting that,

in addition to boosting one's self-esteem, the motivation for women and ethnic minorities to advance the standing of White men might be to avoid the psychic (and actual) punishment that may be meted out to minorities who seek to advance their fellow low-status group members.

A great deal of popular press writing on the glass ceiling suggests it results from minorities preventing their fellow minorities from advancing (Heim & Murphy, 2001). If women and non-Whites would simply stop engaging in "cat fights" and being envious of each other's success, the thinking goes, the glass ceiling would disappear (Eckes, 2002; Epstein, 1980; Tanenbaum, 2002). Indeed, there is a body of research indicating that women are especially competitive with female coworkers (Derks, Van Laar, Ellemers, & de Groot, 2011; Ellemers et al., 2004; Ely, 1994), and that non-Whites are especially competitive with non-White coworkers (Chattopadhyay et al., 2004). However, our results suggest that ethnic minority and women leaders engage in diversity-valuing behavior at their own peril, and that, if they engage in this pursuit, their reputation may be tarnished and their performance ratings may suffer. Thus, our findings help explain why minorities in powerful positions may feel motivated to limit the career progression of talented minorities and women under their command.

Interestingly, we found across both studies that some of the highest competence and performance ratings were awarded to non-White and women leaders who engaged in low levels of diversity-valuing behavior. This finding may explain why women and ethnic minorities may be reticent to advocate for their own groups, as evidenced by the "queen bee" and "crab mentality" effects demonstrated in the tokenism literature (see Sheppard & Aquino, 2013, for a review). One implication of our findings is that engaging in a low level of diversity-valuing behavior may be a critical prerequisite for the upward social mobility of women and non-White leaders. This finding is consistent with evidence showing that minority and women leaders can gain access to the upper echelons of organizations through ingratiating themselves to incumbents (Westphal & Stern, 2006, 2007), downplaying their differences from incumbents (Zhu et al., 2014), and engaging in modern sexism (Watkins et al., 2006).

Strengths, Limitations, and Future Research

This set of studies has several strengths, including experimental and field designs. We also manipulated and subjectively measured diversity-valuing

behavior, which further enhanced confidence in our theoretical predictions (Campbell, Stanley, & Gage, 1963). Thus, we took advantage of the rigor and internal validity of laboratory contexts, and the power and generality of field contexts. In addition, our moderated mediational hypotheses were supported using robust methods (Preacher & Hayes, 2008).

However, like all research, ours has some limitations. One potential problem with Study 1 is that our measure of diversity-valuing behavior focuses somewhat on the leader's underlying value of diversity rather than exclusively focusing on the leader's diversity-valuing behaviors. However, Study 2 was specifically designed to address this concern by manipulating one of the most highly consequential types of leader diversity-valuing behavior. Additionally, we included our Study 1 diversity-valuing behavior measure in Study 2 and found that manipulated diversity-valuing behavior (in Study 2) is associated with higher ratings of diversity-valuing behavior (Study 1 measure). This result bolsters our confidence in our theoretical model. However, future research could examine the extent to which a leader's value of diversity has similar effects to actual diversity-valuing behavior on bosses' stereotypical judgments of the leader as well as performance ratings.

Our Study 1 measure of leader perceived competence may be viewed as a potential weakness of our research. As noted above, in Study 2, we measured this variable in two different ways, using the Study 1 measure as well as an established scale. We found that the Study 1 measure was highly correlated with the established measure. To further enhance confidence in our theoretical predictions, we ran our Study 2 regression analysis replacing the established competence measure with the Study 1 competence measure, and the results were similar to the ones we reported. We recognize that future research could explore how different measures of perceived competence might differentially mediate the joint influence of leader sex/race and diversity-valuing behavior on performance ratings. One other potential issue is that we measured Study 2 perceived competence and performance at the same time and from the same source. However, these two variables were captured from different sources in Study 1, which helps instill confidence in our conceptual model. Future research may benefit by examining whether our conceptual model still holds if a non-White or woman leader engages in diversity-valuing behavior without attracting attention, unlike the leaders in Study 2. Perhaps female or ethnic minority

leaders can protect their competence and performance ratings by covertly, rather than overtly, seeking to hire and promote demographically low-status individuals.⁵

One surprising finding from Study 2 was that non-White and women leaders tended to be rated more positively than their White and male counterparts. In our view, this puzzling finding likely resulted from participant social desirability concerns. For example, White participants tend to give overly positive evaluations of minorities and overly negative evaluations of Whites in order to avoid appearing racist to researchers (Krysan & Couper, 2003) and pollsters (Hopkins, 2008). This “Bradley effect” means that demographically high-status participants tend to tell pollsters and researchers that they favor a minority candidate, although they actually end up voting for a majority candidate (Hopkins, 2008). Thus, applied to our studies, we would expect participants to inflate their evaluations of non-White and female job candidates in Study 2 (to avoid appearing racist and sexist), and these overly positive evaluations would not be reflected in their actual behavior (i.e., Study 1 actual performance evaluations). However, despite this likely social desirability effect, we still observed our hypothesized effects across both studies, which increases our confidence in our findings.

Our research also has several methodological strengths that help bolster confidence in our findings. Our moderated mediation model (Muller, Judd, & Yzerbyt, 2005) linking the interaction of leader diversity-valuing behavior and demographics to performance ratings through the mechanism of stereotypical leader attribute ratings was confirmed using both the Sobel test (MacKinnon et al., 1995) as well as with Preacher and colleague’s bootstrapping methodology (Preacher, Rucker, & Hayes, 2007). Moreover, our full theoretical model was confirmed in a large field sample of 350 high-level executives (CEOs, other C-level leaders, vice presidents, directors, and board-level professionals), as well as a laboratory setting where we were able to manipulate our predictor variables (diversity-valuing behavior and leader demographics). This approach is consistent with the idea of full-cycle research (Chatman & Flynn, 2005), wherein a naturally occurring phenomenon is observed in the field, and then brought into a carefully controlled laboratory setting to verify the causal process and intervening mechanisms. A final strength is that our field study

data were derived from multiple respondents (peer and supervisor ratings) which helps minimize concerns regarding common-method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

Practical Implications

Although one of our contributions is formally introducing the diversity-valuing behavior construct, this construct may do more harm than good for non-White and female leaders. Certainly, one implication of this research is that the entire diversity framing plays a major role in perpetuating the glass ceiling. Engaging in diversity-valuing behavior may be viewed as socially competitive behavior for minorities and women. If organizations (and researchers) would focus on leaders’ homogeneity-valuing behavior rather than their diversity-valuing behavior, the burden of proof would be shifted from those trying to change the glass-ceiling status quo to those trying to maintain it. However, we realize that focusing on leaders’ homogeneity-valuing rather than their diversity-valuing behavior is neither fair nor practically feasible and simply replaces one problem with another. A fairer approach would be to simply measure and reward the degree to which people hire and promote individuals who are demographically dissimilar from themselves. Because White men currently hold a clear numerical majority at the highest organizational levels, rewarding such demographic unselfishness would naturally correct the demographic imbalances throughout organizations, as members of demographic majorities would tend to hire and promote members of demographic minorities. Even if organizations do not change their reward structures, our results suggest the glass ceiling might be weakened if researchers and managers simply stopped focusing on employees’ “diversity-valuing behavior” and instead shifted their focus to “demographic-unselfishness behavior.”

Somewhat counterintuitively, our findings suggest that organizations seeking to advance the standing of minorities and women might consider having a White male spokesperson for the diversity office. Typically, diversity offices are run by ethnic minorities and women, but our results imply that, for maximum legitimacy, organizational leaders might consider changing this. Perhaps because he intuitively sensed our finding, the CEO of United Parcel Service (a White male) serves as the leader of the company’s diversity council because he believes “it makes everyone in the company take diversity issues seriously” (Daft, 2011: 350). Likewise, we found that

⁵ We are grateful to an anonymous reviewer for raising this point.

raters tended to view demographic unselfishness favorably (ethnic minorities and women advocating for hiring a White man) or neutrally (e.g., White men advocating for females and non-Whites).

Ironically, our results suggest that, on balance, the glass ceiling may actually become stronger, rather than weaker, with each ethnic minority or woman leader hired. Because we found that diversity-valuing behavior tends to be a somewhat acceptable behavior for White men, yet an illegitimate behavior for ethnic minorities and women, the latter may be able to advance their own careers to the extent they reinforce the glass ceiling by engaging in a low level of diversity-valuing behavior. Indeed, results from both of our studies show that the highest performance ratings were given to minority leaders who engaged in a low level of diversity-valuing behavior, leading to the problematic conclusion that promoting and championing White men may be a highly beneficial career strategy for ethnic minority and women leaders.

Former Secretary of State Madeline Albright said that there was a special place in hell reserved for women who don't help other women. Our results suggest that powerful women may be damned if they do and damned if they don't leave the ladder up behind them. Indeed, our results suggest that there is a tradeoff for women and non-White leaders in the sense that, for these demographically low-status individuals to succeed, they must avoid advancing the standing of the low-status groups to which they belong. For example, our findings suggest the disturbing implication that low-status demographic groups cannot "lift themselves up by their own bootstraps," but, rather, must rely on members of high-status demographic groups to increase their societal power and status. This implication is troubling, because White men advancing women and non-White leaders may be viewed as condescending behavior by the recipients and reverse discrimination by other White men (e.g., Riley, 2014). That said, there are likely few neat or easy solutions to a problem as intractable and persistent as that of the glass ceiling.

Our theory and results are consistent with anecdotal evidence from two contrasting, powerful women—Marissa Mayer (Yahoo's CEO) and Jill Abramson (fired *New York Times* chief). During her brief tenure at the helm of the *New York Times*, Jill Abramson was an outspoken advocate of women and increased the percentage of female senior editors from 20% to 50% (Sullivan, 2014). In contrast, Marissa Mayer told reporters that she is not a feminist and held the percentage of female top managers

at Yahoo steady at 23% (Isidore, 2014). By distancing herself from her gender, Marissa Mayer's shareholders and board of directors may view her as more competent, and thus as worthy of maintaining her powerful position. Our theoretical model and results indeed suggest that Jill Abramson may have undermined her own career by advocating for diversity at the *New York Times*.

CONCLUSION

Our field study has several strengths, including a large sample spanning 20 industries and 26 job functions, several control variables that help rule out potential alternative explanations, and peer and boss ratings of leaders occupying some of the most powerful positions in the United States. Likewise, our laboratory study has several strengths, including manipulated predictor variables and carefully controlled conditions that rule out potential alternative explanations. Therefore, our findings provide consistent, strong, and compelling support for our theoretical predictions that ethnic minority and female leaders who value diversity will tend to be negatively stereotyped and tend to receive lower performance ratings. Our findings help diagnose one reason why a major organizational problem persists (i.e., the glass ceiling), and thereby provide insight into how researchers and organizations might remedy the problem.

REFERENCES

- Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. 2008. Fundamental dimensions of social judgment. *European Journal of Social Psychology*, 38: 1063–1065.
- Aiken, L. S., & West, S. G. 1991. *Multiple regression: Testing and interpreting interactions*. London, England: Sage.
- Avolio, B. J., Walumbwa, F. O., & Weber, T. J. 2009. Leadership: Current theories, research, and future directions. *Annual Review of Psychology*, 60: 421–449.
- Barnett, W. P., Baron, J. N., & Stuart, T. E. 2000. Avenues of attainment: occupational demography and organizational careers in the California civil service 1. *American Journal of Sociology*, 106: 88–144.
- Baron, R. M., & Kenny, D. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51: 1173–1182.
- Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M. 1977. *Status characteristics and social interaction*. New York, NY: Elsevier.

- Bertrand, M., & Mullainathan, S. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94: 991–1013.
- Bettencourt, B., Charlton, K., Dorr, N., & Hume, D. L. 2001. Status differences and in-group bias: a meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. *Psychological Bulletin*, 127: 520–542.
- Bliese, P. D. 2000. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*: 349–381. San Francisco, CA: Jossey-Bass.
- Brass, D. J. 1985. Men's and women's networks: A study of interaction patterns and influence in an organization. *Academy of Management Journal*, 28: 327–343.
- Brutus, S., Fleenor, J. W., & London, M. 1998. Does 360-degree feedback work in different industries?: A between-industry comparison of the reliability and validity of multi-source performance ratings. *Journal of Management Development*, 17: 177–190.
- Burns, C., Barton, K., & Kerby, S. 2012, July. *The state of diversity in today's workforce: As our nation becomes more diverse so too does our workforce* (Issue brief). Washington, D.C.: Center for American Progress.
- Campbell, D. T., Stanley, J. C., & Gage, N. L. 1963. *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.
- Catalyst, Inc. 2012. *High potentials in the pipeline: Leaders pay it forward*. Available at <http://www.catalyst.org/knowledge/high-potentials-pipeline-leaders-pay-it-forward> (accessed August 9, 2012).
- Catalyst, Inc. 2013. *Women CEOs of the Fortune 1000*. Available at <http://catalyst.org/knowledge/women-ceos-fortune-1000> (accessed January 2013).
- Chatman, J. A., & Flynn, F. J. 2005. Full-cycle micro-organizational behavior research. *Organization Science*, 16: 434–447.
- Chattopadhyay, P., George, E., & Ng, C. K. 2011. An uncertainty reduction model of relational demography. In A. Joshi, H. Liao & J. J. Martocchio (Eds.), *Research in personnel and human resources management*, 30: 219–251. Greenwich, CT: JAI Press.
- Chattopadhyay, P., Tluchowska, M., & George, E. 2004. Identifying the ingroup: A closer look at the influence of demographic dissimilarity on employee social identity. *Academy of Management Review*, 29: 180–202.
- Connelly, B. S., & Ones, D. S. 2010. An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136: 1092–1122.
- Cook, A., & Glass, C. 2014. Above the glass ceiling: When are women and racial/ethnic minorities promoted to CEO? *Strategic Management Journal*, 35: 1080–1089.
- Cuddy, A. J., Fiske, S. T., & Glick, P. 2007. The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92: 631–648.
- Cuddy, A. J., Fiske, S. T., & Glick, P. 2008. Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, 40: 61–149.
- Daft, R. L. 2011. *The leadership experience* (5th ed.). Mason, OH: Cengage.
- Derks, B., Ellemers, N., van Laar, C., & de Groot, K. 2011. Do sexist organizational cultures create the queen bee? *British Journal of Social Psychology*, 50: 519–535.
- Derks, B., Van Laar, C., Ellemers, N., & de Groot, K. 2011. Gender-bias primes elicit queen-bee responses among senior policewomen. *Psychological Science*, 22: 1243–1249.
- Dezsö, C. L., & Ross, D. G. 2012. Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal*, 33: 1072–1089.
- Diversity, Inc. staff. 2012. *Where's the Diversity in Fortune 500 CEOs?* Available at <http://www.diversityinc.com/diversity-facts/wheres-the-diversity-in-fortune-500-ceos/> (accessed September 1, 2014).
- Dorgan, T., & Grieco, M. 1993. Battling against the odds: The emergence of senior women trade unionists. *Industrial Relations Journal*, 24: 151–165.
- Duguid, M. 2011. Female tokens in high-prestige work groups: Catalysts or inhibitors of group diversification? *Organizational Behavior and Human Decision Processes*, 116: 104–115.
- Duguid, M. M., Loyd, D. L., & Tolbert, P. S. 2012. The impact of categorical status, numeric representation, and work group prestige on preference for demographically similar others: A value threat approach. *Organization Science*, 23: 386–401.
- Eagly, A. H., & Karau, S. J. 2002. Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109: 573–598.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. 1992. Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, 111: 543–588.
- Eagly, A. H., & Steffen, V. J. 1984. Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, 46: 735–754.
- Eckes, T. 2002. Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles*, 47: 99–114.

- Ellemers, N., Heuvel, H., Gilder, D., Maass, A., & Bonvini, A. 2004. The underrepresentation of women in science: Differential commitment or the queen bee syndrome? *British Journal of Social Psychology*, 43: 315–333.
- Ellemers, N., Rink, F., Derks, B., & Ryan, M. K. 2012. Women in high places: When and why promoting women into top positions can harm them individually or as a group (and how to prevent this). *Research in Organizational Behavior*, 32: 163–187.
- Elsesser, K. M., & Lever, J. 2011. Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. *Human Relations*, 64: 1555–1578.
- Ely, R. J. 1994. The effects of organizational demographics and social identity on relationships among professional women. *Administrative Science Quarterly*, 39: 203–238.
- Epstein, C. F. 1980. Women's attitudes toward other women: Myths and their consequences. *American Journal of Psychotherapy*, 34: 322–333.
- Fiske, S. T., Cuddy, A. J. C., Glick, P. S., & Xu, J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82: 878–902.
- Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. 1999. (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *The Journal of Social Issues*, 55: 473–489.
- Freeman, D., Aquino, K., & McFerran, B. 2009. Overcoming beneficiary race as an impediment to charitable donations: Social dominance orientation, the experience of moral elevation, and donation behavior. *Personality and Social Psychology Bulletin*, 35: 72–84.
- Gaertner, S. L., Dovidio, J. F., Rust, M. C., Nier, J. A., Banker, B. S., Ward, C. M., Mottola, G. R., & Houlette, M. 1999. Reducing intergroup bias: Elements of intergroup cooperation. *Journal of Personality and Social Psychology*, 76: 388–402.
- Goldin, C., & Rouse, C. 2000. Orchestrating impartiality: The impact of “blind” auditions on female musicians. *The American Economic Review*, 90: 715–741.
- Gutiérrez, A. S., & Unzueta, M. M. 2013. Are admissions decisions based on family ties fairer than those that consider race? Social dominance orientation and attitudes toward legacy vs. affirmative action policies. *Journal of Experimental Social Psychology*, 49: 554–558.
- Hackman, J. R. 1992. Group influences on individuals in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*: 199–267. Palo Alto, CA: Consulting Psychologists Press.
- Hansen, M. T., Ibarra, H., & Peyer, U. 2010. The best-performing CEOs in the world. *Harvard Business Review*, 88: 104–113.
- Harrison, D. A., Kravitz, D. A., Mayer, D. M., Leslie, L. M., & Lev-Arey, D. 2006. Understanding attitudes toward affirmative action programs in employment: Summary and meta-analysis of 35 years of research. *The Journal of Applied Psychology*, 91: 1013–1036.
- Heilman, M. E., Block, C. J., & Stathatos, P. 1997. The affirmative action stigma of incompetence: Effects of performance information ambiguity. *Academy of Management Journal*, 40: 603–625.
- Heim, P., & Murphy, S. A. 2001. *In the company of women*. New York, NY: Penguin.
- Hekman, D. R., Aquino, K., Owens, B. P., Mitchell, T. R., Schilpzand, P., & Leavitt, K. 2010. An examination of whether and how racial and gender biases influence customer satisfaction. *Academy of Management Journal*, 53: 238–264.
- Herring, C. 2009. Does diversity pay? Race, gender, and the business case for diversity. *American Sociological Review*, 74: 208–224.
- Hillman, A. J., Cannella, A. A., & Harris, I. C. 2002. Women and racial minorities in the boardroom: How do directors differ? *Journal of Management*, 28: 747–763.
- Hillman, A. J., Shropshire, C., & Cannella, A. A. 2007. Organizational predictors of women on corporate boards. *Academy of Management Journal*, 50: 941–952.
- Hitt, M. A., & Barr, S. H. 1989. Managerial selection decision models: Examination of configural cue processing. *The Journal of Applied Psychology*, 74: 53–61.
- Hollander, E. P. 1958. Conformity, status, and idiosyncrasy credit. *Psychological Review*, 65: 117–127.
- Hoobler, J. M., Hu, J., & Wilson, M. 2010. Do workers who experience conflict between the work and family domains hit a “glass ceiling?” A meta-analytic examination. *Journal of Vocational Behavior*, 77: 481–494.
- Hoobler, J. M., Wayne, S. J., & Lemmon, G. 2009. Bosses' perceptions of family–work conflict and women's promotability: Glass ceiling effects. *Academy of Management Journal*, 52: 939–957.
- Hopkins, D. J. 2008. *No more Wilder effect, never a Whitman effect: When and why polls mislead about Black and female candidates*. Cambridge, MA: Department of Government, Harvard University. Available at <http://people.iq.harvard.edu/~dhopkins/wilder13.pdf> (accessed October 10, 2008).
- Hultin, M., & Szulkin, R. 1999. Wages and unequal access to organizational power: An empirical test of gender discrimination. *Administrative Science Quarterly*, 44: 453–472.

- Ibarra, H. 1995. Race, opportunity, and diversity of social circles in managerial networks. *Academy of Management Journal*, 38: 673–703.
- Ibarra, H., Carter, N., & Silva, C. 2010. Why men still get more promotions than women. *Harvard Business Review*, 88: 80–85.
- Isidore, C. 2014. Despite Mayer, Yahoo leadership overwhelmingly male. *CNNMoney*, June 18, 2014. <http://money.cnn.com/2014/06/18/technology/yahoo-diversity/> (accessed September 1, 2015).
- James, L., Demaree, R., & Wolf, G. 1984. Estimating within-group interrater reliability with and without response bias. *The Journal of Applied Psychology*, 69: 85–98.
- Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. 2008. The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes*, 106: 39–60.
- Joshi, A. 2014. By whom and when is women's expertise recognized? The interactive effects of gender and education in science and engineering teams. *Administrative Science Quarterly*, 59: 202–239.
- Joshi, A., & Roh, H. 2009. The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*, 52: 599–627.
- Joshi, A., Son, J., & Roh, H. 2015. When can women close the gap? A meta-analytic test of sex differences in performance and rewards. *Academy of Management Journal*, 58: 1516–1545.
- Jost, J. T., Banaji, M. R., & Nosek, B. A. 2004. A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25: 881–919.
- Kanter, R. M. 1977. Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, 82: 965–990.
- Kelly, H. 2015. Google commits \$150 million to diversity. *CNNMoney*, May 7, 2015. <http://money.cnn.com/2015/05/06/technology/google-diversity-plan/> (accessed January 28, 2016).
- Kingstrom, P. O., & Mainstone, L. E. 1985. An investigation of the rater–ratee acquaintance and rater bias. *Academy of Management Journal*, 28: 641–653.
- Kozlowski, S. W. J., & Klein, K. J. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*: 3–90. San Francisco, CA: Jossey-Bass.
- Krysan, M., & Couper, M. P. 2003. Race in the live and the virtual interview: Racial deference, social desirability, and activation effects in attitude surveys. *Social Psychology Quarterly*, 66: 364–383.
- Landy, D., & Sigall, H. 1974. Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology*, 29: 299–304.
- Leslie, L. M., Mayer, D. M., & Kravitz, D. A. 2014. The stigma of affirmative action: A stereotyping-based theory and meta-analytic test of the consequences for performance. *Academy of Management Journal*, 57: 964–989.
- Levi, A. S., & Fried, Y. 2008. Differences between African Americans and Whites in reactions to affirmative action programs in hiring, promotion, training, and layoffs. *The Journal of Applied Psychology*, 93: 1118–1129.
- Lewis, A. C., & Sherman, S. J. 2003. Hiring you makes me look bad: Social-identity based reversals of the ingroup favoritism effect. *Organizational Behavior and Human Decision Processes*, 90: 262–276.
- Lopuch, V. S., & Davis, D. C. 2014. The role and value of diversity to learning organizations and innovation. In N. Erbe (Ed.), *Approaches to managing organizational diversity and innovation*: 213–236. Hershey, PA: IGI Global.
- Lyness, K. S., & Thompson, D. E. 1997. Above the glass ceiling? A comparison of matched samples of female and male executives. *The Journal of Applied Psychology*, 82: 359–375.
- Lyness, K. S., & Thompson, D. E. 2000. Climbing the corporate ladder: Do female and male executives follow the same route? *The Journal of Applied Psychology*, 85: 86–101.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. 2002. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7: 83–104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. 2004. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39: 99–128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. 1995. A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30: 41–62.
- Magee, J. C., & Galinsky, A. D. 2008. Social hierarchy: The self-reinforcing nature of power and status. *The Academy of Management Annals*, 2: 351–398.
- Martinko, M. J., Harvey, P., & Dasborough, M. T. 2011. Attribution theory in the organizational sciences: A case of unrealized potential. *Journal of Organizational Behavior*, 32: 144–149.
- McConahay, J. B. 1986. Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio &

- S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism*: 91–126. Orlando, FL: Academic Press.
- Mendoza, S. L. 2002. *Between the homeland and the diaspora: The politics of theorizing Filipino and Filipino identities*. New York, NY: Routledge.
- Miville, M. L., Gelso, C. J., Pannu, R., Liu, W., Touradji, P., Holloway, P., & Fuertes, J. 1999. Appreciating similarities and valuing differences: The Miville–Guzman universality–diversity scale. *Journal of Counseling Psychology*, 46: 291–307.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. 2007. Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60: 683–729.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. 2005. When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89: 852–863.
- Neeley, T. B. 2013. Language matters: Status loss and achieved status distinctions in global organizations. *Organization Science*, 24: 476–497.
- Ng, T. W., Eby, L. T., Sorensen, K. L., & Feldman, D. C. 2005. Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, 58: 367–408.
- Nishii, L. 2013. The benefits of climate for inclusion for gender diverse groups. *Academy of Management Journal*, 56: 1754–1774.
- O'Brien, K. E., Biga, A., Kessler, S. R., & Allen, T. D. 2010. A meta-analytic investigation of gender differences in mentoring. *Journal of Management*, 36: 537–554.
- O'Brien, L. T., & Major, B. 2005. System-justifying beliefs and psychological well-being: The roles of group status and identity. *Personality and Social Psychology Bulletin*, 31: 1718–1729.
- Oh, I. S., Wang, G., & Mount, M. K. 2011. Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *The Journal of Applied Psychology*, 96: 762–773.
- Park, S. H., & Westphal, J. D. 2013. Social discrimination in the corporate elite how status affects the propensity for minority CEOs to receive blame for low firm performance. *Administrative Science Quarterly*, 58: 542–586.
- Parker, C. P., Baltes, B. B., & Christiansen, N. D. 1997. Support for affirmative action, justice perceptions, and work attitudes: A study of gender and racial–ethnic group differences. *The Journal of Applied Psychology*, 82: 376–389.
- Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. 2014. Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *The Journal of Applied Psychology*, 99: 1129–1145.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, 88: 879–903.
- Powell, G. N., & Butterfield, D. A. 1994. Investigating the “glass ceiling” phenomenon: An empirical study of actual promotions to top management. *Academy of Management Journal*, 37: 68–86.
- Preacher, K. J., & Hayes, A. F. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40: 879–891.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. 2007. Assessing moderated mediation hypotheses: Strategies, methods, and prescriptions. *Multivariate Behavioral Research*, 42: 185–227.
- Ragins, B. R., & Scandura, T. A. 1999. Burden or blessing? Expected costs and benefits of being a mentor. *Journal of Organizational Behavior*, 20: 493–509.
- Ragins, B. R., Townsend, B., & Mattis, M. 1998. Gender gap in the executive suite: CEOs and female executives report on breaking the glass ceiling. *The Academy of Management Executive*, 12: 28–42.
- Riley, J. L. 2014. Opinion: Why liberals should stop trying to “help” black Americans. *New York Post*, June 28, 2014. <http://nypost.com/2014/06/28/how-liberals-make-it-harder-for-blacks-to-succeed/> (accessed September 1, 2014).
- Rosette, A. S., Leonardelli, G. J., & Phillips, K. W. 2008. The White standard: Racial bias in leader categorization. *The Journal of Applied Psychology*, 93: 758–777.
- Ryan, M. K., & Haslam, S. A. 2005. The glass cliff: Evidence that women are over-represented in precarious leadership positions. *British Journal of Management*, 16: 81–90.
- Sackett, P. R., & DuBois, C. L. 1991. Rater–ratee race effects on performance evaluation: Challenging meta-analytic conclusions. *The Journal of Applied Psychology*, 76: 873–877.
- Sadri, G., Weber, T. J., & Gentry, W. A. 2011. Empathic emotion and leadership performance: An empirical analysis across 38 countries. *The Leadership Quarterly*, 22: 818–830.
- Schaefer, R. T. 1996. Education and prejudice. *The Sociological Quarterly*, 37: 1–16.
- Schneider, B., Ehrhart, M. G., Mayer, D. M., Saltz, J. L., & Niles-Jolly, K. 2005. Understanding organization–customer links in service settings. *Academy of Management Journal*, 48: 1017–1032.
- Sears, D. O. 1988. Symbolic racism. In P. A. Katz & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy*: 53–84. New York, NY: Plenum Press.

- Sheppard, L. D., & Aquino, K. 2013. Much ado about nothing? Observers' problematization of women's same-sex conflict at work. *The Academy of Management Perspectives*, 27: 52–62.
- Sidanius, J., Pratto, F., & Mitchell, M. 1994. In-group identification, social dominance orientation, and differential intergroup social allocation. *The Journal of Social Psychology*, 134: 151–167.
- Spector, P. A., & Brannick, M. T. 2011. Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, 14: 287–305.
- Staines, G., Tavis, C., & Jayaratne, T. E. 1973. The queen bee syndrome. *Psychology Today*, 7: 55–60.
- Sullivan, M. 2014. The *Times*, from the top: Looking ahead. *New York Times*, January 12, 2014: SR12. Available at http://www.nytimes.com/2014/01/12/public-editor/the-times-from-the-top-looking-ahead.html?_r=0 (accessed September 1, 2014).
- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. 1995. Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68: 199–214.
- Swim, J. K., & Miller, D. L. 1999. White guilt: Its antecedents and consequences for attitudes toward affirmative action. *Personality and Social Psychology Bulletin*, 25: 500–514.
- Tanenbaum, L. 2002. *Catfight: Women & competition*. New York, NY: Seven Stories Press.
- The Economist*. 2008. Briefing: Black America—Nearer to overcoming. *The Economist*, May 8: 37–40.
- Tosi, H. L., Werner, S., Katz, J. P., & Gomez-Mejia, L. R. 2000. How much does performance matter? A meta-analysis of CEO pay studies. *Journal of Management*, 26: 301–339.
- Turban, D. B., & Jones, A. P. 1988. Supervisor–subordinate similarity: Types, effects, and mechanisms. *The Journal of Applied Psychology*, 73: 228–234.
- Unzueta, M. M., & Knowles, E. D. 2014. The “business case” for diversity may not by itself make the strongest case for diversity. In *Diversity ideologies in organizations*: 257–267. Taylor and Francis.
- Van Knippenberg, D., & Schippers, M. C. 2007. Work group diversity. *Annual Review of Psychology*, 58: 515–541.
- Walker, H. J., Field, H. S., Bernerth, J. B., & Becton, J. B. 2012. Diversity cues on recruitment websites: Investigating the effects of job seekers' information processing. *The Journal of Applied Psychology*, 97: 214–224.
- Watkins, M. B., Kaplan, S., Brief, A. P., Shull, A., Dietz, J., Mansfield, M.-T., & Cohen, R. 2006. Does it pay to be a sexist? The relationship between modern sexism and career outcomes. *Journal of Vocational Behavior*, 69: 524–537.
- Wayne, J. H., & Casper, W. J. 2012. Why does firm reputation in human resource policies influence college students? The mechanisms underlying job pursuit intentions. *Human Resource Management*, 51: 121–142.
- Wenneras, C., & Wold, A. 2001. Nepotism and sexism in peer-review. In M. Wyer (Ed.), *Women, science, and technology: A reader in feminist science studies*: 46–52. New York, NY: Routledge.
- Westphal, J. D., & Stern, I. 2006. The other pathway to the boardroom: Interpersonal influence behavior as a substitute for elite credentials and majority status in obtaining board appointments. *Administrative Science Quarterly*, 51: 169–204.
- Westphal, J. D., & Stern, I. 2007. Flattery will get you everywhere (especially if you are a male Caucasian): How ingratiation, boardroom behavior, and demographic minority status affect additional board appointments at U.S. companies. *Academy of Management Journal*, 50: 267–288.
- Zhu, D. H., Shen, W., & Hillman, A. J. 2014. Recategorization into the in-group: The appointment of demographically different new directors and their subsequent positions on corporate boards. *Administrative Science Quarterly*, 59: 240–270.
- Zimmerman, R. D., Triana, M. D. C., & Barrick, M. R. 2010. Predictive criterion-related validity of observer ratings of personality and job-related competencies using multiple raters and multiple performance criteria. *Human Performance*, 23: 361–378.
- Zoogah, D. B. 2010. Why should I be left behind? Employees' perceived relative deprivation and participation in development activities. *The Journal of Applied Psychology*, 95: 159–173.
- Zweigenhaft, R. L., & Domhoff, G. W. 2006. *Diversity in the power elite: How it happened, why it matters*. Lanham, MD: Rowman & Littlefield.
- Zweigenhaft, R. L., & Domhoff, G. W. 2011. *The new CEOs: Women, African American, Latino, and Asian American leaders of Fortune 500 companies*. Lanham, MD: Rowman & Littlefield.



David R. Hekman (david.hekman@colorado.edu) earned his PhD at the University of Washington's Foster School of Business, and is an associate professor of organizational leadership and analytics at the University of Colorado's Leeds School of Business. He is focused on improving organizational health by examining remedies for pervasive workplace racial and gender biases, sources of professional workers' motivation, and sources and outcomes of virtuous leadership. His previous work has been published in *Academy of Management Journal*, *Journal of Applied Psychology*, and *Organization Science*, among others.

Stefanie K. Johnson (stefanie.johnson@colorado.edu) earned her PhD at Rice University, and is an assistant professor of organizational leadership and analytics at the University of Colorado's Leeds School of Business. Her research focuses on issues of leadership and diversity. She has published more than 40 journal articles and book chapters in outlets such as *Journal of Applied Psychology*, *Organizational Behavior and Human Decision Processes*, and *The Leadership Quarterly*.

Maw-Der Foo (foo@nus.edu.sg) is an associate professor at the Division of Engineering & Technology Management, National University of Singapore. His research is at the intersection of management and entrepreneurship, and explores attitudinal, affective, and cognitive factors that

result in the success and well-being of entrepreneurs. His work has appeared in journals including the *Academy of Management Journal*, *Entrepreneurship Theory & Practice*, *Journal of Applied Psychology*, *Journal of Business Venturing*, *Management Science*, and *Strategic Management Journal*.

Wei Yang (wei.yang@utexas.edu) is a doctoral student in management at the McCombs School of Business, University of Texas at Austin. Her research interests include technology management, technological change, and the antecedents of innovation, including diversity in organizations.



Copyright of Academy of Management Journal is the property of Academy of Management and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.