We attempt to make sense of ongoing gender disparities in the upper ranks of organizations by examining gender bias in leaders’ assessments of managers’ derailment potential. In a large managerial sample (Study 1: \( N \sim 12,500 \)), we found that ineffective interpersonal behaviors were slightly less frequent among female managers but slightly more damaging to women than men when present. Evidence of bias was not found in performance evaluations but emerged when leaders were asked about derailment potential in the future. We replicated this pattern of effects in a second large managerial sample (Study 2: \( N \sim 35,500 \)) and in two experimental studies (Studies 3 and 4) in which gender and interpersonal behaviors were manipulated. In Study 4, we also showed that when supervisors believe that a manager might derail in the future, they tend to withdraw mentoring support and sponsorship, which are especially critical for women’s career advancement. Our research highlights the importance of leaders’ perceptions of derailment potential—which differ from evaluations of performance or promotability—both because they appear to be subject to stereotype-based gender bias and because
they have important implications for the mentoring and sponsorship that male and female managers receive.

There is a long tradition of beginning discussions of gender at work by grappling with these facts: Women continue to be underrepresented in upper management and are rare at the officer and director levels of major corporations (Catalyst, 2013a, 2013b). Yet, women are equally represented in the workforce (Catalyst, 2014b), earn the majority of the bachelor’s and graduate degrees awarded in the United States (U.S. Department of Education, 2012), and hold more than half of management, professional and related positions (Catalyst, 2014a): The challenge is not getting into organizations but getting to the top. Given that gaps exist in the upward mobility of men and women even when controlling for potential differences such as leadership style (Eagly & Johnson, 1990), clustering in gendered occupations (Morrison & Von Glinow, 1990), and geographic mobility (Stroh, Brett, & Reilly, 1992), the question remains as to what prevents women from advancing to the upper echelons of organizational leadership.

One possible answer might be that women’s performance is not evaluated as positively as men’s. Yet, meta-analyses show that this is not universally the case (Bowen, Swim, & Jacobs, 2000; Eagly, Karau, & Makhijani, 1995; Joshi, Son, & Roh, 2015; Roth, Purvis, & Bobko, 2012). There is a catch, however: these studies also demonstrate that the same performance evaluations may be differentially predictive of advancement for men and women. Roth et al. (2012) found that although women’s job performance evaluations were higher than men’s, men were rated higher on promotion potential, and Joshi et al. (2015) determined that despite a lack of significant differences in men and women’s performance evaluations, substantial gaps persisted in promotions and other rewards. Lyness and Heilman (2006) confirmed that women appear to be required to perform at a higher standard than men to be promoted, suggesting that bias stemming from gender stereotypes may still play a role in hindering women’s promotions.

Whether or not gender affects workplace evaluations has been the subject of some debate in the literature. For example, Landy (2008) asserted that promotion decisions are structured assessments characterized by clear and individuating information, which tends to override stereotype activation; thus, gender bias should not play a role in such decisions and the best man or woman should win. Whether this is true or not, we argue that the current debate fails to capture the whole picture. The potential effects of stereotype-based bias on promotions cannot be understood solely by examining performance ratings or promotion decisions themselves: by the time supervisors evaluate a manager for promotion, stereotype-based
bias may already have taken a subtle toll on women’s potential by influencing earlier and more ambiguous judgments that ultimately affect promotability.

Thus, we focus on informal evaluations that have received little attention in the management literature, but that have powerful implications for whether women rise or flatline in organizations: bosses’ assessments of derailment potential. Managers are said to have “derailed” when they fail to achieve anticipated career outcomes and are involuntarily demoted, fired, or plateaued without reaching the level expected of them (Lombardo, Ruderman, & McCauley, 1988). We propose that ineffective interpersonal behaviors, considered a prime warning sign of derailment (Leslie & Van Velsor, 1996), are disproportionately damaging for women because they violate female gender stereotypes of communality, thus activating shifting standards of evaluation for men and women. We also demonstrate that when bosses believe an employee might derail, they withdraw mentoring of the type most important to women’s advancement.

We make a unique empirical contribution by testing these propositions in two large archival datasets, each containing thousands of multisource feedback reports, as well as in two experiments in which we manipulated gender and interpersonal behaviors to examine their links to derailment potential as well as the association between perceptions of derailment potential and mentoring. Researchers have long proposed that ambiguous evaluations, such as potential for future derailment, leave room for stereotype-based bias (e.g., Heilman, 2001), but there is a lack of large-scale field studies demonstrating this. We advance theory on gender bias in organizations by connecting these biased evaluations to withdrawal of supervisory mentoring and sponsorship, which are especially critical to women’s career advancement. The subtle biases we study may not affect formal performance assessments, but they play an important role in career success, especially for women.

**Gender Stereotypes and Bias in Organizations**

Two major theoretical perspectives have guided recent research on gender stereotypes as a barrier to advancement. The first is Eagly and Karau’s (2002) role congruity theory. This theory posits that role-based expectations for men and leaders are aligned: Both are characterized by agentic behavior. In contrast, role-based expectations for women, characterized by communal behavior, are misaligned with the leader role, producing a “think manager, think male” phenomenon (Heilman, Block, Martell, & Simon, 1989; Schein, 1973). This results in the catch-22 situation frequently described in management literature (Phelan, Moss-Racusin, & Rudman, 2008; Rudman & Glick, 2001), wherein communal women are
not perceived as a good fit for high-responsibility management positions (Heilman, 1983; Lyness & Heilman, 2006), but women who are not perceived as communal, even when they are effective, are evaluated harshly (Heilman & Okimoto, 2007; 2008; Rudman & Phelan, 2008).

The second theory, Biernat’s shifting standards model (Biernat, 2003), posits that different standards may be used in evaluating the same behavior, depending on whether it comes from a man or a woman. Biernat and Kobynowicz (1999) explained that the reference points for subjective evaluations of traits and behaviors are derived from gender-based stereotypes: “The terms good or bad, tall or short, emotional or unemotional . . . denote different things when they are applied to men than to women: They imply, for example, ‘emotional, for a woman,’ or ‘short, for a man,’ even when these tags are not explicit” (p. 82). Recent experimental research demonstrated that identical interpersonal behaviors at work are judged differently based on the actor’s gender; observers viewed the same work conflicts as more problematic when they occurred between women than between men (Sheppard & Aquino, 2013).

In combination, these theories suggest that there will be few differences in others’ evaluations of men and women when they exhibit behaviors that are expected in the workplace (i.e., effective interpersonal behaviors). Rather, differences in their evaluations will emerge when stereotypes are violated. When male and female managers exhibit good interpersonal behaviors, no stereotypes are activated; such behaviors are expected from managers and will lead to similar, positive evaluations for men and women. It is only when role-based stereotypes are activated or violated that shifting standards of evaluation begin to emerge. When men and women engage in ineffective interpersonal behaviors, men will tend to be excused (“his behaviors were not that bad for a man”) and women will tend to be penalized (“her behaviors were really bad for a woman”).

In this study, we focus on managers’ interpersonal behaviors not only because they are deeply related to gender stereotypes about communality but because previous research indicates that interpersonal skills are a critical component of managerial success and a core predictor of managerial derailment. Interpersonal skills are more important than business or strategic skills at all organizational levels, and their importance increases at high levels of the organization (Mumford, Campion, & Morgeson, 2007). Conversely, ineffective interpersonal behaviors, which have been operationalized as having poor working relationships (Leslie & Van Velsor, 1996), experiencing problems with interpersonal relationships (Cullen, Gentry, & Yammarino, 2015), exhibiting dysfunctional interpersonal tendencies (Carson et al., 2012), or being perceived by colleagues as insensitive, cold, or arrogant (Kovach, 1986), are among the best predictors of managerial derailment.
We argue then that ineffective interpersonal behaviors will affect leaders’ perceptions of managerial derailment potential differently for men and women. Due to gender-based expectations for communal behavior among women, ineffective interpersonal behaviors are more damaging for women than for men. Women who exhibit these behaviors violate gender role expectations for communality (i.e., role congruity theory; Eagly & Karau, 2002), leading their behaviors to be judged more harshly than would be the same behavior exhibited by a man (i.e., shifting standards of evaluation; Biernat, 2003). In addition, we argue that gender stereotypes do not affect all judgments equally in the workplace. When supervisors face structured, relatively public decisions guided by clear criteria and concrete information, as in formal performance evaluations, these factors tend to override the influence of stereotypes (Landy, 2008). This accounts for empirical evidence indicating that performance evaluations in the 21st century may not be subject to much bias against women (e.g., Joshi et al., 2015; Roth et al., 2012), but gender bias is more likely to color subtle judgments that bosses make about the future. Stereotypical information is more influential in biasing decisions when evaluation criteria are ambiguous, when evaluation processes are unstructured (Heilman, 2001; Heilman, Block, & Stathatos, 1997) and when the decision maker is powerful relative to the target (Fiske, 1993; Goodwin, Gubin, Fiske, & Yzerbyt, 2000). One such bias-prone judgment in organizations is leaders’ assessments of managers’ potential for derailment in the future.

**Derailment**

Derailment is defined as occurring “when a manager who was expected to go higher in the organization and who was judged to have the ability to do so is fired, demoted, or plateaued below expected levels of achievement” (Lombardo & McCauley, 1988, p. 1). What distinguishes derailment from more widely researched management constructs such as poor performance, lack of promotion, or involuntary turnover is the element of disappointed expectations. Actual derailment can be thought of as the deviation between what leaders expected for an employee and what that employee actually achieves in terms of career outcomes. Previous work on derailment has emphasized that derailment is involuntary; intentionally opting out of advancement is not derailment (Lombardo & Eichinger, 1989; Lombardo et al., 1988).

Derailment may unfold over the course of years, as it becomes clear that a manager once thought to be a rising star may not have the competencies needed to meet the growing demands associated with higher level positions (Lombardo & Eichinger, 1989; Zhang, Leslie, & Hannum, 2013). For this reason, we are consistent with the literature in focusing on
the initial seeds of derailment: leaders’ assessments of the likelihood that a manager will derail in the future. Studies of the Pygmalion and Golem effects in organizations have shown that leaders’ expectations of their subordinates’ outcomes shape their behavior toward subordinates and elicit corresponding behavioral responses from the subordinates themselves, creating self-fulfilling prophecies (McNatt, 2000). In this way, bosses’ perceptions of managers’ future derailment potential may impact actual derailment in both direct and indirect ways.

Derailment and derailment potential have received relatively limited attention as a focus of organizational research, but they are heated topics in organizations. Popular management books advise ambitious managers that awareness and careful control of their self-sabotaging behavioral tendencies are necessary to avoid being derailed (e.g., de Haan & Kasozi, 2014; Dotlich & Cairo, 2003; Irwin, 2012). Even leaders who do not perceive themselves to be at risk have a strong interest in trying to predict who else in their organizations might go off the rails, given that estimates of the costs of a failed senior manager or executive can run into the millions of dollars (Hogan, Hogan, & Kaiser, 2010). Unsurprisingly, an Internet search turns up a plethora of practitioner-focused guides to “spotting the warning signs,” with interpersonal problems topping the lists (e.g., Brodnitzki, 2014; Priolo, 2011; Stebbins, 2014). Content analysis on interviews with senior leaders in the United States and Europe also identified relationship problems as a primary cause of managerial derailment (Leslie & Van Velsor, 1996; Morrison, White, & Van Velsor, 1987; Van Velsor & Leslie, 1995), a finding that has been supported in a variety of research studies (e.g., Gentry, Mondore, & Cox, 2007; Gentry & Shanock, 2008; Irani Williams, Campbell, McCartney, & Gooding, 2013). The literature also suggests that interpersonal behaviors are particularly important in shaping assessments of derailment potential for female managers (Gentry, Clark, Young, Cullen, & Zimmerman, 2015; Lombardo & Eichinger, 1989).

Unlike performance evaluations, which Landy (2008) argued tend to be based heavily on the behavior and output of individuals, assessments of a manager’s potential for future derailment involve more ambiguous judgments. When executives evaluate the future potential of men and women, stereotype-based biases come into play because projections into the future are uncertain. Perceptions regarding derailment are especially subject to stereotype-based biases because they are not recorded formally as part of managers’ personnel files in the way that performance evaluations and promotions are, making them less available for scrutiny. If women in a department, unit, or organization systematically receive lower performance evaluations than men, or are not promoted as frequently, it is possible for others in the organization to observe these practices; they show up
in organizational reports and, at least in many large corporations, attract attention. But informal judgments, such as perceptions of derailment risk, are off the record and may be private or even unconscious, making them difficult to identify and address.

Exhibiting ineffective interpersonal behaviors at work could flag any employee as a derailment risk, but such behaviors may be especially problematic for women. Thus:

*Hypothesis 1:* There is a positive association between ineffective interpersonal behaviors and derailment potential.

*Hypothesis 2:* The positive association between ineffective interpersonal behaviors and derailment potential is stronger for women than men, and these differences increase as levels of ineffective interpersonal behaviors increase.

**Derailment, Mentoring, and Sponsorship**

Perceiving that a subordinate is at risk of derailment might not change a supervisor’s behavior in overt ways (e.g., giving a poor performance evaluation). Indeed, interviews with executives suggest that they often do not take direct action, instead hoping that managers will correct their own courses (Brodnitzki, 2014). But perceptions of derailment potential may have subtle effects on opportunities for growth and advancement in organizations. As managers are groomed for future positions, they are given challenging assignments and other platforms for increased visibility within the organization as well as introductions to powerful others (Ragins, Townsend, & Mattis, 1998). Because beneficial resources accrue to employees who are held in positive regard by their supervisors (Graen & Uhl-Bien, 1995), managers who are thought to be at risk for derailment may not be given these key opportunities. Thus, when it is time for promotion decisions, high- and low-derailment potential candidates might have similar formal performance assessments, but very different experiences, social networks, and informal cachet.

Exposure to influential people in organizations is among the most important support functions for subordinates’ advancement: The more central people are in dominant organizational networks, the more likely they are to be promoted (Brass, 1985). Previous research suggests that exposure is disproportionately critical to career advancement for women, who may struggle to break into powerful male-oriented networks on their own (Bevelander & Page, 2011; Ibarra, 1992; Ibarra, 1993; Timberlake, 2005). People tend to form more and stronger network ties to those who are part of their own identity groups (Ibarra, 1992), making it difficult for women to develop the informal connections they need to advance because men
at the top of organizations may see them as illegitimate or untrustworthy “outsiders” (Burt, 1998). Powerful sponsors within the organization can help women by providing them with challenging assignments that can get them noticed, as well as by brokering introductions—enabling women to “borrow” the sponsor’s social capital or “reflected power” (Kanter, 1976) and the accompanying transferal of trust and legitimacy (Bevelander & Page, 2011; Burt, 1998). Burt (1998) found that women who made use of such borrowed capital were promoted earlier, whereas promotions lagged for women who took an entrepreneurial approach to constructing their own social networks. Timberlake (2005) argued that “[f]or women, networks dominated by strong ties with support from strategic sponsors results in the greatest probability of promotion, especially to the highest executive ranks” (pp. 39–40). Although men can also benefit from this support (Burt, 1998), they are less dependent on it for entry into dominant coalition networks (Brass, 1985).

Although this highly visible form of support is valuable to subordinates, especially women, it also poses risks for supervisors. Recent research has begun to draw distinctions between mentoring behaviors, those in which mentors advise and develop protégés in the context of their dyadic relationship, and sponsorship behaviors, those in which mentors “leverage their own power and reputational capital” (Foust-Cummings, Dinolfo, & Kohler, 2011, p. 1) and “help protégés develop skills” (p. 5) to actively drive protégés’ advancement. Mentoring behaviors, such as coaching, providing psychosocial support, and steering a subordinate away from potentially damaging contact with others in the organization (Ibarra, Carter, & Silva, 2010; Ragins & Cotton, 1999), are relatively low risk because they can occur privately, unseen by others. But by engaging in sponsorship behaviors, such as exposing a subordinate to greater visibility in the organization, giving a subordinate a challenging assignment on which failure would reflect poorly on the supervisor, or using personal influence to help a subordinate advance, a supervisor places his or her own reputation at stake (Foust-Cummings et al., 2011; Hewlett, Peraino, Sherbin, & Sumberg, 2010) and stands to lose social capital if the subordinate fails to live up to expectations. Thus, these high-risk forms of support are likely to be the ones most sharply withdrawn if a subordinate appears to be at risk for derailment.

**Hypothesis 3:** There is a negative association between derailment potential and mentoring and sponsorship behaviors.

**Hypothesis 4:** The negative association between derailment potential and mentoring/sponsorship is stronger for sponsorship behaviors (sponsor, expose, and challenge) than for mentoring behaviors (coach, support, and protect).
Overview of the Present Research

The central premises of our theory are that assessments of derailment potential, which represent expectations that leaders have about the future potential of a manager, are (a) subject to gender bias because they involve ambiguous future projections, and (b) important determinants of leaders’ decisions about whether and how much to invest in managers’ futures. We examine leaders’ assessments of the probability that managers will derail in the future because these evaluations represent an early signal of derailment. We focus on ineffective interpersonal behaviors because they are associated with gender stereotypes and have been directly linked to derailment. We assess the impact of gender stereotype-based bias on managerial perceptions of derailment potential in four studies. In Study 1, we use an archival dataset comprising 12,503 managers and their bosses to test our hypothesis that women are perceived as more likely to derail than men when both exhibit ineffective interpersonal behaviors. Study 2 replicates this core finding with a second archival dataset containing 35,583 managers and their supervisors. Two additional studies utilize experimental designs, providing support for the field results. Study 3 addresses the uniqueness of the derailment potential construct, and Study 4 links perceptions of derailment potential to withdrawal of supervisory sponsorship and mentoring.

Study 1: Primary Field Study

Participants

Archival data on 12,503 managers from a wide variety of organizations and industries were obtained via a multisource leadership instrument administered as part of external (conducted outside of the organization) leadership development programs between 2004 and 2008. Online surveys were administered to participants, their direct reports, their peers, and their bosses. Leadership data were included as part of the feedback process, but performance and derailment potential ratings were collected for research purposes only. Participants were primarily male (67.8%) and Caucasian (69.7%). They were 42.82 years old on average (SD = 6.94) and had 17.23 years (SD = 2.74) of education, with 89.2% holding at least a bachelor’s degree. Most participants in these leadership programs were sponsored by their employers. Our participants were broadly representative of managers employed by organizations with the resources to invest in developing leaders; managers in small businesses were underrepresented. Participants represented five managerial levels: first-level supervisors (1.5%), middle managers (18.9%),
upper–middle managers (47.6%), executives (28.9%), and top managers (e.g., CFO, CEO; 3.2%). They also represented a wide range of industries, including education, government, aerospace, manufacturing, banking, and pharmaceuticals.

**Measures**

**Ineffective interpersonal behaviors.** Our measure of interpersonal behaviors was drawn from the derailment section of Benchmarks® (CCL, 2004), a 360-degree feedback instrument used primarily for leadership development (Lombardo & McCauley, 1994; Lombardo, McCauley, McDonald-Mann, & Leslie, 1999; McCall, Lombardo, & Morrison, 1988). This instrument is well validated (see Leslie & Fleenor, 1998; McCauley & Lombardo, 1990) and has been used in numerous research studies (e.g., Graves, Ohlott, & Ruderman, 2007; Lyness & Judiesch, 2008). The 10-item scale is focused on ineffective interpersonal behaviors and includes items such as “Is not adaptable to many different types of people” and “Orders people around rather than working to get them on board.” Ratings were provided by direct reports and peers using a five-point scale that ranged from 1 = strongly disagree to 5 = strongly agree.

On average, each manager was evaluated by 7.63 raters ($SD = 2.18$, median = 7). We computed $r^*_{WG(I)}$ (Lindell, Brandt, & Whitney, 1999) using a uniform null distribution. A median score of .95 and a mean score of .88 indicate strong agreement among raters (LeBreton & Senter, 2008), justifying aggregation of ratings to a single score for each manager ($\alpha = .94$).

**Derailment potential.** Derailment potential was assessed by managers’ bosses with a single item included in a section of Benchmarks that is for research only and is not used for managerial feedback. Bosses indicated the likelihood that managers would “derail in the next 5 years as a result of his/her behavior or actions,” using a five-point scale: 1 = not at all likely, 2 = not very likely, 3 = somewhat likely, 4 = likely, and 5 = almost certain. This single-item indicator of derailment potential is positively correlated with involuntary turnover (being fired) in subsequent years ($r = .19, p < .05; Carson et al., 2012$). A high correlation between derailment potential and involuntary turnover is not expected because termination represents only the extreme outcome of derailment potential.

**Controls.** We obtained data on managers’ job performance, level of management, functional area, industry, sector (public or private), and number of direct reports. We controlled for functional area, industry, and sector because women who work in male-dominated industries may be held to different standards than women working in female-dominated industries (Eagly, Karau, & Makhijani, 1995). We controlled for level of
management because effective interpersonal behaviors may be especially advantageous for top-level leaders (Rosette & Tost, 2010). We controlled for job performance because effective interpersonal behaviors may be less associated with derailment for individuals who are high performers (McCartney & Campbell, 2006). Managers’ performance was rated by bosses with four items (Cronbach’s alpha = .91; e.g., “How would you rate this person’s performance in his/her present job?”; “Where would you place this person as leader relative to other leaders inside and outside your organization?”); responses ranged from 1 = among the worst to 5 = among the best. Level of management, functional area, industry, and sector were reported by participants and were dummy coded. We used four dummy variables for level of management, with “top managers” serving as the reference category. There were 25 functional areas (e.g., sales, marketing, and engineering), leading to 24 dummy variables, with “other” as the reference category. Industry had 11 categories (e.g., manufacturing, health, and government agency), so 10 dummy variables were created, with “other” serving as the reference category. Finally, sector was a dichotomous variable, with private sector = 1 and public sector = 0.

Measurement models. To test whether our measures of ineffective interpersonal behaviors, derailment potential, and performance captured distinct constructs, we used confirmatory factor analysis (CFA; Mplus 7; Muthén & Muthén, 2012) to test a model wherein items loaded onto their respective latent factors, with correlated latent factors and uncorrelated residuals. Because derailment potential was measured by a single item, we set its factor loading to one and its error variance to zero. We also examined two two-factor models. In the first, we combined derailment potential with performance to form a single factor; in the second, we combined derailment potential with ineffective interpersonal behaviors.

Results indicated that our three-factor measurement model fit the data well ($\chi^2 = 8,283.91$, df = 88, $p < .001$, comparative fit index [CFI] = .96, root mean square error of approximation [RMSEA] = .09, standardized root mean square residual [SRMR] = .03), with all items having strong and statistically significant loadings on their designated factors; the standardized factor loadings ranged from .83 to .87 for performance (mean = .85) and from .79 to .93 for ineffective interpersonal behaviors (mean = .87). Chi-square difference tests revealed a significant decrement in fit for the two two-factor models ($\Delta \chi^2 = 356.26$, $\Delta df = 1$, $p < .001$ for the two-factor model combining derailment potential and performance; and $\Delta \chi^2 = 4126.07$, $\Delta df = 1$, $p < .001$ for the two-factor model combining derailment potential and ineffective interpersonal behaviors), and fit statistics were weaker for the two-factor models. Given these results, we
treated ineffective interpersonal behaviors, performance, and derailment potential as distinct constructs.

*Measurement invariance tests.* To test whether our measure of ineffective interpersonal behaviors and performance functioned similarly for men and women, we conducted a CFA-based multigroup measurement invariance test. Following Facteau and Craig (2001), we combined ineffective interpersonal behavior ratings from direct reports and peers and fit a CFA model, loading the 10 ineffective interpersonal behavior items onto a latent factor with men and women in the same model. We used a sandwich estimator to obtain robust standard errors in Mplus 7 software (i.e., including the syntax TYPE = COMPLEX under the ANALYSIS command) to account for nonindependence of observations (multiple ratings for the same manager; Rogers, 1993). Results show a good fit for this model ($\chi^2 = 12,752.69, df = 80, p < .001, CFI = .97, TLI = .96, RMSEA = .06, SRMR = .03$); all items loaded significantly on the latent factor for both men (standardized factor loadings ranged from .71 to .85, mean = .79) and women (standardized factor loadings ranged from .72 to .87, mean = .79). We repeated this analysis for performance and found similarly good fit ($\chi^2 = 414.16, df = 8, p < .001, CFI = .99, TLI = .98, RMSEA = .09, SRMR = .02$) and significant factor loadings for men (standardized factor loadings ranged from .83 to .87, mean = .85) and women (standardized factor loadings ranged from .83 to .88, mean = .85).

Next, we specified a model in which factor loadings were constrained to be equivalent for men and women. Results showed that the invariant model fit the data well for ineffective interpersonal behaviors ($\chi^2 = 13,209.23, df = 89, p < .001, CFI = .97, TLI = .97, RMSEA = .06, SRMR = .03$) and performance ($\chi^2 = 420.37, df = 11, p < .001, CFI = .99, TLI = .99, RMSEA = .08, SRMR = .02$). For performance, there was no significant difference between the baseline model and the model wherein men and women were constrained to be equal ($\Delta \chi^2 = 6.21, \Delta df = 3, ns$). For interpersonal behaviors, Satorra’s (2000) $\chi^2$ difference test showed a significant increase in $\chi^2$ in the constrained model compared to the baseline model (Satorra–Bentler scaled $\Delta \chi^2 = 486.28, \Delta df = 9, p < .01$), which indicates a worse fit. Because the $\chi^2$ statistic is very sensitive in large samples such as ours (Facteau & Craig, 2001), it can lead to rejection of well-fitting models. Thus, incremental fit indices, such as CFI and RMSEA, should be used along with $\chi^2$ to assess best model fit. These indices were equally good in the baseline and constrained models.

Considering the evidence in its entirety, these results suggest both congeneric and tau-equivalent invariance for the interpersonal behaviors scale and ratings of job performance; measurement error is unlikely to account for differences in observed scores for men and women.
Results

Descriptive statistics. Tables 1 and 2 present descriptive statistics and correlations. Although these descriptive statistics are informative as to the nature of our data, they cannot speak to the issue of gender bias. Men had significantly \((d = .04, p < .001)\) more ineffective interpersonal behaviors than women did, though this difference was small. Consistent with existing literature (Lyness & Heilman, 2006; Roth et al., 2012), we found no differences between men and women on either performance or derailment potential. As hypothesized, ineffective interpersonal behaviors were moderately and significantly correlated with derailment potential (Hypothesis 1: \(r = .30, p < .001\)) and with job performance \((r = -.26, p < .001)\).

Ineffective interpersonal behaviors and derailment potential for men and women. We examined the associations between ineffective interpersonal behaviors, gender, and boss-rated derailment potential and performance (Table 3). We found that ineffective interpersonal behaviors were negatively associated with performance \((\beta = -.22, p < .001)\) even after controlling for organizational level, job function, industry, sector, and number of raters. Gender was not associated with performance \((\beta = .00, ns)\) nor was the interaction term between gender and ineffective interpersonal behaviors significant \((\beta = -.02, ns)\) in predicting performance.

Next, we tested the association between gender, ineffective interpersonal behaviors, and derailment potential. As expected (Hypothesis 1), bosses’ ratings of the likelihood of derailment were higher for managers with more ineffective interpersonal behaviors \((\beta = .26, p < .001)\), and this association remained even after controlling for performance \((\beta = .15, p < .001)\). In general, women and men were viewed by their bosses as equally likely to derail \((\beta = -.01, ns)\), but women were more likely than men to be at risk for derailment as they exhibited increasingly higher levels of ineffective interpersonal behaviors \((\beta = .02, p < .01)\), supporting Hypothesis 2. Simple slope analysis showed a slightly stronger association between ineffective interpersonal behaviors and derailment potential for women \((\beta = .17, p < .001)\) than for men \((\beta = .15, p < .001)\).

To further illustrate the nature of this interaction, we conducted simple slope tests of gender effect on derailment at multiple values of ineffective interpersonal behaviors. Results showed that at the lowest level of ineffective interpersonal behaviors (scale value = 1), women were rated less likely to derail than men \((\beta = -.08, p < .01)\), an effect we did not hypothesize. This difference became nonsignificant as ineffective interpersonal behaviors increased to the scale value of 2 \((\beta = .01, ns)\). As the level of ineffective interpersonal behaviors increased, the trend flipped such that women were rated more likely to derail than men at the scale value of 3 \((\beta = .09, p < .05)\); this gap continued to widen as ineffective
<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall sample</th>
<th></th>
<th>Men</th>
<th></th>
<th>Women</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>F</td>
</tr>
<tr>
<td>Study 1: Ineffective interpersonal behaviors</td>
<td>1.79</td>
<td>.52</td>
<td>1.80</td>
<td>.52</td>
<td>1.76</td>
<td>.51</td>
<td>20.33***</td>
</tr>
<tr>
<td>Performance</td>
<td>3.99</td>
<td>.81</td>
<td>3.99</td>
<td>.81</td>
<td>3.98</td>
<td>.81</td>
<td>.31</td>
</tr>
<tr>
<td>Derailment potential</td>
<td>1.85</td>
<td>.97</td>
<td>1.86</td>
<td>.96</td>
<td>1.83</td>
<td>.98</td>
<td>2.38</td>
</tr>
<tr>
<td>Study 2: Effective interpersonal behaviors</td>
<td>4.04</td>
<td>.56</td>
<td>4.02</td>
<td>.56</td>
<td>4.09</td>
<td>.57</td>
<td>104.88***</td>
</tr>
<tr>
<td>Derailment potential</td>
<td>2.12</td>
<td>.90</td>
<td>2.13</td>
<td>.89</td>
<td>2.09</td>
<td>.91</td>
<td>12.81***</td>
</tr>
</tbody>
</table>

Note. Study 1: N = 12,503 (8,479 men; 4,024 women).
Study 2: N = 35,583 (23,995 men; 11,588 women).

***p < .001.
interpersonal behaviors increased ($\beta = .17, p < .05$ for scale value 4; $\beta = .25, p < .01$ for scale value 5). As expected, gender bias is most evident at the highest levels of ineffective interpersonal behaviors where the largest violation of female stereotypes surrounding communal behavior occurs.

**Practical significance.** To quantify the practical significance of our results, we conducted a binary logistic regression that provided us with odds ratios estimating the likelihood of derailment for a man and woman with the same interpersonal behaviors. We dichotomized the derailment variable into low risk (“not at all likely” and “not very likely” were coded 0) and high risk (“likely” and “almost certain” were coded 1). We then ran a logistic regression with our block of control variables along with gender, ineffective interpersonal behaviors, and the interaction between them. The odds ratio for the interaction term was 1.38 ($p < .05$). We then computed the log odds of derailment for men and women at high and low levels of ineffective interpersonal behaviors (Cody & Smith, 1997), then converted log odds into odds, and then probabilities. When ineffective interpersonal behaviors were low (scale scores of 1 and 2), the probability of derailment was similar for men and women (6.2% for men and 6.3% for women), but when ineffective interpersonal behaviors were high (scale scores of 4 and 5), the likelihood of derailment was 17% higher for women (60.3% for men and 77.1% for women).

**Post hoc analyses.** To determine if the gender interactions we found varied by the context in which managers worked, as suggested by a recent meta-analysis of experimental studies (Koch, D’Mello, & Sackett, 2015), we tested three three-way interactions. The first two were the interactions between gender, ineffective interpersonal behaviors, and the gendered nature of (a) the job function and (b) industry. The gendered natures of both function and industry were computed as the percentage of
<table>
<thead>
<tr>
<th></th>
<th>Study 1</th>
<th>Study 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance</td>
<td>Derailment potential</td>
<td>Performance</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>4.16***</td>
<td>1.79***</td>
<td>1.90***</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy variables for organizational level ($R^2$)</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Dummy variables for organizational functional area ($R^2$)</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Dummy variables for industrial type ($R^2$)</td>
<td>.06</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Dummy variable for organizational sector</td>
<td>.03*</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>Number of raters</td>
<td>.04***</td>
<td>-.02*</td>
<td>.00</td>
</tr>
<tr>
<td>Performance</td>
<td>-</td>
<td>-</td>
<td>-53***</td>
</tr>
<tr>
<td><strong>Study variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ineffective interpersonal behaviors</td>
<td>-.22***</td>
<td>.26***</td>
<td>.15***</td>
</tr>
<tr>
<td>Effective interpersonal behaviors</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gender</td>
<td>.00</td>
<td>.00</td>
<td>-.01</td>
</tr>
<tr>
<td>Ineffective interpersonal behaviors × Gender</td>
<td>-.02</td>
<td>.03**</td>
<td>.02**</td>
</tr>
<tr>
<td>Effective interpersonal behaviors × Gender</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.11</td>
<td>.10</td>
<td>.35</td>
</tr>
<tr>
<td>$\Delta R^2$</td>
<td>.000</td>
<td>.001**</td>
<td>.001**</td>
</tr>
</tbody>
</table>

**Note.** Coefficients, unless otherwise labeled (e.g., $R^2$), are standardized regression coefficients. In Study 1, we measured ineffective interpersonal behaviors; in Study 2, we measured effective interpersonal behaviors. $\Delta R^2$ represents change in $R^2$ when the interaction term between interpersonal behaviors and gender was added into the regression.

*p < .05, **p < .01, ***p < .001.
women in our dataset, which ranged from 0.0% to 61.5% for job function and from 15.7% to 56.4% for industry. We validated the gendered percentages in our data by comparison to archival data on engineering (e.g., 11.3% women in engineering in our data compared to approximately 14.8% women in engineering [National Science Foundation, 2015], and 61.5% women in human resources in our data compared to 71% women in human resources management [Goudreau, 2011]). Third, we examined the interaction between gender, ineffective interpersonal behaviors, and organizational level ($1 = \text{first-level supervisors}$ to $5 = \text{top managers}$).

None of these three-way interactions was significant ($\beta = .01, ns, \beta = .00, ns, \text{and } \beta = .02, ns$, for job function, industry type, and organizational level, respectively). Moreover, the two-way interactions between ineffective interpersonal behaviors and gender remained significant in each case ($\beta_s = .02, ps < .05$ for all three regressions).

Because the justification for some of our controls also suggested moderation, we examined several additional interactions. First, we examined the interaction between performance and ineffective interpersonal behaviors in predicting derailment potential. We found significant effects for the interaction term ($\beta = −.06, p < .01$), such that the positive association between ineffective interpersonal behaviors and derailment potential was stronger for managers with worse job performance ($\beta = .20, p < .01$) than those with better job performance ($\beta = .09, p < .01$). Based on these results, we reestimated our regressions with both interactions (i.e., gender $\times$ ineffective interpersonal behaviors and job performance $\times$ ineffective interpersonal behaviors); both were significant ($\beta = .02, p < .01$ for gender; and $\beta = −.06, p < .01$, for job performance). We also tested the three-way interaction among gender, job performance, and ineffective interpersonal behaviors in influencing derailment potential; it was not significant ($\beta = −.02, ns$). Finally, we examined the interaction between managerial level and ineffective interpersonal behaviors in predicting derailment potential. No significant interaction was found ($\beta = .01, ns$).

Brief discussion. Study 1 results show that the same behaviors from a man and woman lead to a different evaluation of the behaviors, which ultimately lead to different assessments of derailment potential. These findings are consistent with predictions of both role congruity theory and the shifting standards model: women pay a higher premium than men with the same level of ineffective interpersonal behaviors. This bias toward women is not revealed in performance evaluations, which focus on current behaviors and outcomes, but emerges when bosses are asked about derailment potential in the future, resulting in 17% greater likelihood of derailment for women when ineffective interpersonal behaviors are high.

In Study 2, we retest our central hypothesis—that the association between interpersonal behaviors and derailment potential is different for
men and women—in another large managerial sample. This constructive replication provides a rigorous test of our hypotheses because Study 2 data include a measure of effective, rather than ineffective, interpersonal behaviors. We expect to find smaller gender bias in this sample because we are comparing men and women who vary on the extent to which they exhibit effective interpersonal behaviors; failing to exhibit effective interpersonal behaviors is a less egregious violation of gender stereotypes for communality than is exhibiting a high level of ineffective interpersonal behaviors. We also seek to determine if women are advantaged, compared to men, when they exhibit the highest level of effective interpersonal behaviors.

Study 2: Field Sample Replication

Participants

Archival data from multirater feedback reports obtained in similar circumstances to those in Study 1 were obtained from an international management consulting firm. In this sample, reports of interpersonal relationships and derailment were provided by bosses. Data from 35,583 participants were used for this analysis. They were primarily male (67.4%) and Caucasian (67.6%), were 41.2 years old on average (SD = 7.87), and 84.1% held at least a bachelor’s degree. As in Study 1, managers were drawn from all management levels (5.4% supervisor, 43.5% first-line management, 35.7% middle management, 13.6% executive management, and 1.8% top management [e.g., CFO, CEO]).

Measures

Effective interpersonal behaviors. We extracted a measure of effective interpersonal behaviors from the PROFILOR® for managers (PDI Ninth House, 2004), a multisource tool used for leadership assessment and development; it has also been used in research (see Johnson & Ferstl, 1999; Scullen, Mount, & Goff, 2000). We identified 12 items that dealt specifically with building or maintaining positive interpersonal relationships, including items such as “Compromises to build give-and-take relationships with others,” “Treats people with respect,” and “Develops effective working relationships with peers.” Exploratory factor analysis revealed one factor with an eigenvalue over 1.00, suggesting unidimensionality for the 12-item measure (Cronbach’s alpha = .92). Bosses indicated the extent to which each manager displayed these behaviors, using a five-point scale ranging from 1 = not at all to 5 = a very great extent.

Derailment potential. The current sample included a single item assessing derailment potential: “What is this person’s risk of experiencing career difficulty due to factors under his/her personal control?” Bosses
responded using a five-point rating scale: 1 = little to no risk, 2 = slight risk, 3 = moderate risk, 4 = high risk, and 5 = very high risk. Evidence of construct validity can be drawn from this item’s moderate association with assessments of promotability (r = −.33, p < .001; N = 14,461 in a separate dataset with primarily midlevel managers).

**Controls.** As in Study 1, we used managerial level and industry as controls, coded in the same way, but the Study 2 data did not include section, job performance, job function, or sector.

**Results**

Table 1 shows that women exhibited more effective interpersonal behaviors than men (d = .07, p < .001) and were also perceived by their bosses as less likely to derail (d = −.04, p < .001). As expected, effective interpersonal behaviors were strongly and negatively correlated with derailment potential (Table 2; above the diagonal; r = −.45, p < .001). The interaction found in Study 1 was replicated. Effective interpersonal behaviors were more strongly related to derailment potential for women than for men (Table 3: β = −.02, p < .05); simple slopes for men and women, respectively: β = −.44, p < .001 and β = −.46, p < .001.

As in Study 1, we conducted simple slope tests of gender effects on derailment at various values of interpersonal behavior. At lower scale values of effective interpersonal behaviors, women were rated more likely to derail than men (β = .15, p < .01 at scale value of 1; β = .11, p < .01 at scale value of 2). The discrepancy became smaller but still significant as effective interpersonal behaviors increased (β = .07, p < .01 at scale value of 3; β = .02, p < .05 at scale value of 4) and finally became nonsignificant when effective interpersonal behaviors reached the highest possible value (β = −.02, ns at scale value of 5). As hypothesized, this pattern suggests gender bias in ratings of derailment potential that begin to emerge as women violate stereotype-based expectations for communal behavior. In this study, we found no evidence of an advantage for women at any level of effective interpersonal behaviors.

**Practical significance.** We followed Study 1’s procedures, first dichotomizing the derailment variable (0 = little to no risk or slight risk and 1 = high risk or very high risk) so that we could run a logistic regression, and then computing odds ratios and probabilities of derailment for men and women at various levels of interpersonal behavior. The odds ratio for the interaction between gender and effective interpersonal behaviors was significant (odds ratio = .67, p < .001). When men and women exhibited highly effective interpersonal behaviors (scale scores of 4 and 5), their probabilities of derailment were similar (2.7% for men and 2.5% for women), but when effective interpersonal behaviors were low (scale
scores of 1 and 2), women were 3.5% more likely to derail than men (93.1% for men and 96.6% for women).

Study 1 and Study 2: Brief Discussion

Both Studies 1 and 2 suggest gender bias in evaluations of derailment potential when stereotypes for communal behavior are violated, calling into question Landy’s (2008) assertion that gender bias is unlikely to occur in organizations. Our results are more consistent with claims that—especially for ambiguous judgments—gender bias does occur, even in natural work settings (Leslie, King, Bradley, & Hebl, 2008).

In Study 1 and Study 2, we obtained archival data from two organizations that used different items to measure derailment potential. The Study 1 item was a direct assessment of the likelihood of future derailment, whereas the Study 2 item was a more indirect measure; experiencing career difficulty is clearly included in the derailment construct, but the Study 2 measure is broad enough to include short-term or temporary career difficulties that might not ultimately lead to derailment. In Study 3, we examine the validity of these measures.

Study 3: Validation

Participants and Procedure

We recruited 160 U.S. managers (mean age = 44.0; 58.8% male; mean managerial experience = 4.0 years) via The Study Response Project, a service that facilitates recruitment of adult research participants and has been used in previous management research (e.g., Piccolo & Colquitt, 2006). Participants were paid $20.00 to review a report described as “the 360 evaluation form for one middle manager” as part of a study evaluating multisource feedback as a managerial development tool. Participants were randomly assigned to review a report for a man (Jason Mayors) or woman (Jennifer Mayors) in the form of a full-color PDF file, modeled on a Benchmarks® (CCL, 2004) report.

Reports for Jennifer and Jason were identical except for first names, gendered pronouns (e.g., he vs. she), and photographs on the cover of the report. Jennifer and Jason are among the top 20 names given to babies in the United States in the 1970s and 1980s, a time when our models would have been born. We portrayed them using stock photos of a man and woman of around 30 years of age in an office context. Equality in attractiveness of these photos was established by ratings of 387 undergraduate business students (not participants in this study) who rated the photos (1 = very attractive, 3 = average, and 5 = very unattractive; mean = 2.51 for each, \( t = -.09, ns \)).
In the report, we used a gender-neutral job title for the hypothetical middle manager: Level III Administrative Manager in Purchasing (see Heilman & Chen, 2005). The feedback report included the company name (InformaTech), background information about the manager (education, prior work experience, employee ID, name, and a color photograph), information about raters (two bosses, six direct reports, five peers, and two others), the norm group (middle managers), and what was assessed (job behaviors). The report included ratings on core job tasks (e.g., analyzing inventory, establishing suppliers, and managing staff), as well as behaviors associated with stalled careers (e.g., difficulty adapting, failure to meet business objectives, and relationship behaviors). In the report, Jennifer (Jason) was rated slightly above the mean of middle managers on all core job tasks except for managing staff, on which she (he) was rated a bit lower than the norm group. Jennifer (Jason) scored more favorably than other managers on failing to meet business objectives and less favorably on interpersonal behaviors. In summary, the feedback report portrayed Jennifer (Jason) as a high-performing manager who had some difficulty with relationships. We included a section labeled “Highlights” to reinforce key information, along with rater comments drawn from a pool of actual comments made by raters of managers similar to those in Study 1 and Study 2.

**Measures**

**Derailment potential.** We used a two-item composite of the derailment items from Study 1 and Study 2 (“How likely is it that Jennifer [Jason] will experience career difficulty due to factors under her [his] personal control?” and “What is the likelihood that Jennifer [Jason] will derail in the next 5 years as a result of her [his] actions or behaviors?”; response options ranged from 1 = not at all likely to 3 = somewhat likely to 5 = almost certain).

**Promotability.** We included two items to measure promotability: “How likely is it that Jennifer (Jason) will be considered for promotion?” and “How likely is it that Jennifer (Jason) will be promoted to an upper management position in the next several years?”

**Performance.** We used two items: “How would you rate Jennifer’s (Jason’s) performance over the past year?”; “How would you rate Jennifer’s (Jason’s) performance in core job tasks?” Response options ranged from 1 = poor to 3 = average to 5 = excellent.

**Results**

We used exploratory factor analysis (EFA) to examine the factor structure of the six items, using principle component extraction and direct
oblimin (oblique) rotation. Results showed three factors with eigenvalues greater than 1; the scree plot also suggested a three-factor structure. The rotated pattern matrix revealed three factors with the two derailment potential items as the first factor (.81 and .86), the two promotability items as the second factor (.90 and .89), and the two performance items as the third factor (.73 and .89). A follow-up CFA showed that the three-factor model fit the data well ($\chi^2 = 11.46$, $df = 6$, $ns$, $CFI = .97$, $TLI = .94$, $RMSEA = .08$, $SRMR = .04$). All items significantly loaded on their respective latent factors, with standardized factor loadings from .44 to .95. Derailment potential was not correlated with performance ($r = .09$, $ns$) and was moderately correlated with promotability ($r = .29$, $p < .05$). Performance and promotability were strongly correlated ($r = .49$, $p < .01$). We also tested a one-factor model, loading all items onto a single latent factor; this model fit the data poorly ($\chi^2 = 61.89$, $df = 9$, $p < .01$, $CFI = .75$, $RMSEA = .19$, $SRMR = .10$; $\Delta \chi^2 = 50.41$, $\Delta df = 3$, $p < .01$).

To further test our hypotheses, we also compared means for each of the three variables (derailment, promotability, and performance) across conditions. Because the reports for Jennifer and Jason were exact duplicates except for gender, a mean difference in ratings between conditions is indicative of gender effects similar to those in Studies 1 and 2. Results reveal significant differences in derailment ratings; Jennifer was viewed as more likely to derail than Jason (mean derailment ratings 3.47 and 3.20, respectively; $t = -2.147$, $p < .05$). No significant differences were found for performance (3.84 for Jennifer and 3.95 for Jason; $t = 1.1$, $ns$) or promotability (3.38 for Jennifer and 3.39 for Jason, $t = .04$, $ns$).

Results of Study 3 provide preliminary evidence of the uniqueness of the derailment potential construct and the validity of the composite measure. In addition, they replicate the results of Study 1 by revealing gender bias only for derailment potential and not for performance. In Study 4, we conduct a final experiment to begin the process of examining outcomes associated with leaders’ perceptions of derailment potential.

**Study 4: Derailment and Mentoring**

**Participants and Procedure**

We recruited a sample of 180 master workers from Amazon’s Mechanical Turk. The sample was 53.9% female and 40 years old on average ($SD = 11.84$), with 4.63 years of management experience ($SD = 6.24$). As in Study 3, participants reviewed a report for Jennifer or Jason and answered a set of questions based on it; they were paid $5.00.

*Experimental manipulation.* Slightly adapting our Study 3 manipulation, we created four versions of the report: one for Jason with low levels
of ineffective interpersonal behaviors, one for Jason with high levels of ineffective interpersonal behaviors, one for Jennifer with low levels of ineffective interpersonal behaviors, and one for Jennifer with high levels of ineffective interpersonal behaviors. Therefore, this study has a 2 (gender: male vs. female) × 2 (ineffective interpersonal behaviors: low vs. high) between-subject design.

Low versus high ineffective interpersonal behaviors were manipulated with the manager’s average scores from raters, displayed on a graph and reinforced in the “Highlights” summary. In both conditions, the manager was rated slightly more favorably than the mean (compared to middle managers) on the majority of core job tasks as well as on meeting business objectives. What we manipulated were scores on elements associated with interpersonal relationships. In the low ineffective interpersonal behaviors condition, Jennifer and Jason scored at the average of managers on managing staff and relationship problems. In the high ineffective interpersonal behaviors condition, they scored higher than average on problems with interpersonal relationships. The report also included qualitative comments from raters regarding the manager’s areas of excellence and opportunities for development. A sample developmental comment in the high ineffective interpersonal behaviors report stated, “She [He] is very task-focused and that can sometimes make her [him] abrupt with others.” In contrast, developmental feedback in the low ineffective interpersonal behaviors condition did not mention the manager’s interpersonal behaviors, instead including comments such as “Continue to grow and gain more responsibility.”

We adapted the 10-item scale of ineffective interpersonal behaviors from Study 1 as a manipulation check (e.g., the original item “has an unresolved interpersonal conflict with boss” became “has unresolved interpersonal conflicts with others”; α = .95). Participants provided higher ratings on this scale in the high ineffective interpersonal behaviors (M = 2.90) than in the low ineffective interpersonal behaviors condition (M = 1.62; t(df = 178) = 13.14, p < .001), providing support for the efficacy of our manipulation.

Measures

Derailment potential. We used the two item scales from Study 3.

Mentoring/sponsorship. We used Ragins and McFarlin’s (1990) Mentor Role Instrument to assess the extent to which mentoring would be provided to each hypothetical manager. Dimensions (three items each) include sponsor (“I would use my influence to support her [his] advancement in our organization”), coach (“I would suggest specific strategies for achieving her [his] career aspirations”), protect (“I would shield her [him]
from damaging contact with important people in the organization”), challenge (“I would assign her [him] tasks that push her [him] into developing new skills”), and expose (“I would help her [him] to be more visible in the organization”). In addition, we modified three social support items in the Mentor Role Instrument slightly to fit our context and included them as a sixth dimension of mentoring (e.g., “I would offer her [him] support and encouragement”). The response options of these items ranged from 1 = strongly disagree to 5 = strongly agree.

Results

Preliminary analysis. Mean, standard deviations, reliabilities, and correlations are presented in Table 4. Derailment potential was negatively and significantly correlated with all mentoring dimensions ($r$s $= −.21 \sim −.66$, $p$s $< .01$), providing preliminary support for the hypothesized relationships between derailment potential and mentoring behaviors.

We conducted a CFA to examine whether our measures of derailment potential and mentoring captured distinct constructs. A seven-factor CFA was tested by loading items on their respective latent factors, with correlated latent factors and uncorrelated residuals. Results revealed a good fit for the data ($\chi^2 = 207.93$, $df = 149$, $p < .01$, CFI = .97, RMSEA = .05, SRMR = .04), with all items having strong and significant loadings on their designated factors; the standardized factor loadings ranged from .73 to .94. An alternative three-factor model was fit with mentoring items loaded onto two latent factors (mentoring factor with coach, support, protect; and sponsorship factor with sponsor, expose, and challenge) to see if mentoring and sponsorship represented distinct higher order factors. This model did not fit the data well ($\chi^2 = 778.85$, $df = 167$, $p < .01$, CFI = .73, RMSEA = .14, SRMR = .10; $\Delta \chi^2 = 570.92$, $\Delta df = 18$, $p < .01$). Modification indices suggested that this model fit poorly because of differential associations between derailment and specific behaviors within the sponsorship and mentoring factors. Thus, we proceeded to test our substantive hypotheses with all six mentoring/sponsorship dimensions.

Hypothesis testing. To test our theoretical model, we fit a structural equation model with all hypothesized relationships included. In this model, we first created latent variables for derailment potential and each of the mentoring dimensions. Then, we specified the effects of gender ($0 = male$, $1 = female$), ineffective interpersonal behaviors ($0 = low$, $1 = high$), and their interaction on derailment potential and the six types of mentoring/sponsorship. Finally, we specified the effects of derailment potential on the mentoring/sponsorship dimensions.

Unstandardized coefficient estimates are presented in Table 5, and key findings are summarized in Figure 1. This structural model fit the data
TABLE 4  
Study 4: Intercorrelations

<table>
<thead>
<tr>
<th>Study variables</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Rater gender</td>
<td>.54</td>
<td>.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Rater age</td>
<td>39.55</td>
<td>11.84</td>
<td>.11</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Rater managerial experience (years)</td>
<td>4.63</td>
<td>6.24</td>
<td>−.02</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gender manipulation</td>
<td>.50</td>
<td>.50</td>
<td>−.06</td>
<td>.02</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Ineffective interpersonal behaviors</td>
<td>.50</td>
<td>.50</td>
<td>−.15</td>
<td>.14</td>
<td></td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Derailment potential</td>
<td>2.13</td>
<td>.81</td>
<td>−.10</td>
<td>.00</td>
<td></td>
<td>.06</td>
<td>−.02</td>
<td></td>
<td></td>
<td>.49</td>
<td>.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Sponsor</td>
<td>3.76</td>
<td>.85</td>
<td>.08</td>
<td>.09</td>
<td>−.02</td>
<td>.04</td>
<td>−.32</td>
<td>−.66</td>
<td>.49</td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Expose</td>
<td>3.90</td>
<td>.70</td>
<td>.01</td>
<td>.09</td>
<td>−.02</td>
<td>.04</td>
<td>−.21</td>
<td>−.43</td>
<td>.63</td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Challenge</td>
<td>4.36</td>
<td>.65</td>
<td>.06</td>
<td>.14</td>
<td>−.01</td>
<td>−.01</td>
<td>−.40</td>
<td>.50</td>
<td>.49</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Coach</td>
<td>4.22</td>
<td>.69</td>
<td>.05</td>
<td>.13</td>
<td>−.02</td>
<td>−.02</td>
<td>.00</td>
<td>−.34</td>
<td>.48</td>
<td>.53</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Support</td>
<td>4.28</td>
<td>.65</td>
<td>.00</td>
<td>.04</td>
<td>−.02</td>
<td>.02</td>
<td>−.14</td>
<td>−.39</td>
<td>.50</td>
<td>.59</td>
<td>.58</td>
<td>.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Protect</td>
<td>3.09</td>
<td>.89</td>
<td>.03</td>
<td>.05</td>
<td>−.10</td>
<td>.03</td>
<td>−.10</td>
<td>−.21</td>
<td>.42</td>
<td>.52</td>
<td>.38</td>
<td>.45</td>
<td>.25</td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 180; rater gender was coded as 0 = male and 1 = female; gender manipulation was coded as 0 = male and 1 = female; ineffective interpersonal behaviors was coded as 0 = low levels of ineffective interpersonal behaviors and 1 = high levels of ineffective interpersonal behaviors. Scale reliabilities are on the diagonal.  
*p < .01, **p < .001.
### Study 4: Unstandardized Estimates of the Hypothesized Model

<table>
<thead>
<tr>
<th>Study variables</th>
<th>Derailment potential</th>
<th>Sponsor</th>
<th>Expose</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>−.01</td>
<td>.10</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td>Ineffective interpersonal behaviors</td>
<td>.73***</td>
<td>.10</td>
<td>.14</td>
<td>.12</td>
</tr>
<tr>
<td>Gender × Interpersonal behaviors</td>
<td>.40</td>
<td>.19</td>
<td>−.23</td>
<td>.19</td>
</tr>
<tr>
<td>Derailment potential</td>
<td>−.88***</td>
<td>.11</td>
<td>−.53***</td>
<td>.10</td>
</tr>
<tr>
<td>Residual variances</td>
<td>.32***</td>
<td>.05</td>
<td>.27**</td>
<td>.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study Variables</th>
<th>Coach</th>
<th>Support</th>
<th>Protect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>−.03</td>
<td>.09</td>
<td>.00</td>
</tr>
<tr>
<td>Ineffective interpersonal behaviors</td>
<td>.37**</td>
<td>.12</td>
<td>.17</td>
</tr>
<tr>
<td>Gender × Interpersonal behaviors</td>
<td>.02</td>
<td>.19</td>
<td>−.01</td>
</tr>
<tr>
<td>Derailment potential</td>
<td>−.51***</td>
<td>.10</td>
<td>−.44***</td>
</tr>
<tr>
<td>Residual variances</td>
<td>.27***</td>
<td>.05</td>
<td>.21***</td>
</tr>
</tbody>
</table>

*Note.* N = 180; Gender manipulation was coded as 0 = male and 1 = female; Ineffective interpersonal behaviors was coded as 0 = low levels of ineffective interpersonal behaviors and 1 = high levels of ineffective interpersonal behaviors. Factor loadings of latent factors on their respective items and covariances among latent factors are not reported. Interested readers can request the information from the corresponding author.

*p < .05, **p < .01, ***p < .001.
Figure 1: Study 4: Unstandardized Estimates of the Hypothesized Model.

Note. N = 180. For the sake of brevity, control effects are not included but are reported in Table 5. Factor loadings of latent factors on their respective items and covariances among latent factors are not depicted as well. Interested readers can request the information from the corresponding author.

*p < .05, ** p < .001.

well, $\chi^2 = 239.83$, df = 188, $p < .01$, CFI = .98, RMSEA = .04, and SRMR = .04. Predictors included in the model accounted for 31.0% of the total variance in derailment potential; variance explained in the mentoring dimensions ranged from 55.5% for the sponsor dimension to 7.6% for the protect dimension.

As shown in Table 5 and Figure 1, ineffective interpersonal behaviors were positively related to derailment potential ($\gamma = .73$, $p < .001$), supporting Hypothesis 1. Derailment potential was rated as more likely in the high ineffective interpersonal behaviors condition ($M = 2.52$) than in the low ineffective interpersonal behaviors condition ($M = 1.74$).

The interaction between ineffective interpersonal behaviors and gender was positively related to derailment potential ($\gamma = .40$, $p < .05$), as well. Simple slope tests showed that the association between the ineffective interpersonal behaviors and derailment potential was significantly stronger for Jennifer ($\gamma = .93$, $p < .001$) than for Jason ($\gamma = .53$, $p < .001$), $d = .40$, $p < .05$, providing additional support for Hypothesis 2.

Ratings of derailment potential were negatively related to all six mentoring dimensions: $\gamma = -.88$, $p < .001$ for support; $\gamma = -.53$, $p < .001$ for expose; $\gamma = -.51$, $p < .001$ for challenge; $\gamma = -.51$, $p < .001$ for coach; $\gamma = -.44$, $p < .001$ for support; and $\gamma = -.30$, $p < .05$ for protect, supporting Hypothesis 3.
Next, we tested the moderated mediation implied by our model. With a 2,000-replication bias-corrected bootstrap procedure, results showed that the indirect effects of ineffective interpersonal behaviors on all six mentoring dimensions via derailment potential were significant; see Table 6, column 5. These results show that ineffective interpersonal behaviors led to participants’ higher ratings of derailment potential, which, in turn, lowered intentions to engage in future mentoring of all types. As hypothesized, the largest effects were for the higher risk sponsorship dimensions, with smaller effects for the lower risk mentoring dimensions. These indirect effects of ineffective interpersonal behaviors on all six mentoring dimensions via derailment potential were significant in both conditions (see columns 2 and 3 of Table 6), but the effects were significantly larger for Jennifer than for Jason (Table 6, column 4). Thus, ineffective interpersonal behaviors led to greater withdrawal of mentoring behaviors across all dimensions (via derailment potential perceptions) when the manager was female.

To further examine Hypothesis 4, we compared the magnitude of the associations between derailment potential and the six mentoring dimensions. The notion that perceptions of derailment potential will lead to greater withdrawal of sponsorship behaviors than mentoring behaviors was generally supported: The effects of derailment potential on sponsor ($\gamma = -0.88, p < .001$), expose ($\gamma = -0.53, p < .001$), and challenge

<table>
<thead>
<tr>
<th>Outcome variables</th>
<th>Jennifer condition</th>
<th>Jason condition</th>
<th>Jennifer and Jason conditions</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sponsor</td>
<td>-.82</td>
<td>-.46</td>
<td>-.35</td>
<td>-.64</td>
</tr>
<tr>
<td></td>
<td>[-1.14, -.52]</td>
<td>[-.77, -.22]</td>
<td>[.03, .74]</td>
<td>[-.90, -.43]</td>
</tr>
<tr>
<td>Expose</td>
<td>-.49</td>
<td>-.33</td>
<td>-.16</td>
<td>-.38</td>
</tr>
<tr>
<td></td>
<td>[-.82, -.26]</td>
<td>[-.66, -.11]</td>
<td>[.06, .32]</td>
<td>[-.65, -.21]</td>
</tr>
<tr>
<td>Challenge</td>
<td>-.48</td>
<td>-.27</td>
<td>-.21</td>
<td>-.37</td>
</tr>
<tr>
<td></td>
<td>[-.77, -.21]</td>
<td>[-.53, -.11]</td>
<td>[.02, .50]</td>
<td>[-.60, -.18]</td>
</tr>
<tr>
<td>Coach</td>
<td>-.47</td>
<td>-.27</td>
<td>-.20</td>
<td>-.37</td>
</tr>
<tr>
<td></td>
<td>[-.85, -.21]</td>
<td>[-.55, -.10]</td>
<td>[.03, .53]</td>
<td>[-.66, -.17]</td>
</tr>
<tr>
<td>Support</td>
<td>-.41</td>
<td>-.23</td>
<td>-.18</td>
<td>-.32</td>
</tr>
<tr>
<td></td>
<td>[-.79, -.18]</td>
<td>[-.52, -.08]</td>
<td>[.03, .43]</td>
<td>[-.62, -.14]</td>
</tr>
<tr>
<td>Protect</td>
<td>-.28</td>
<td>-.16</td>
<td>-.12</td>
<td>-.22</td>
</tr>
<tr>
<td></td>
<td>[-.54, -.04]</td>
<td>[-.35, -.03]</td>
<td>[.01, .35]</td>
<td>[-.42, -.03]</td>
</tr>
</tbody>
</table>

*Note. N = 180. All confidence intervals were generated with a 2,000-replication bias-corrected bootstrap procedure.*
BONO ET AL. 29

\( \gamma = -0.51, p < .001 \) were larger than the effects of derailment on coach
\( \gamma = -0.51, p < .001 \), support \( \gamma = -0.44, p < .001 \), and protect \( \gamma = -0.30, p < .05 \). We also conducted parameter comparison tests in Mplus software by including the “Model Constraint” command, which allows direct comparison of parameters in the model. Results showed that the effect of derailment potential on sponsorship was significantly stronger than its effects of any of the other types of mentoring behaviors \( d = .35, p < .01 \) for expose; \( d = .37, p < .01 \) for challenge; \( d = .37, p < .01 \) for coach; \( d = .44, p < .001 \) for support; and \( d = .58, p < .001 \) for protect), but none of the other associations between derailment potential and mentoring behaviors differed significantly. Thus, Hypothesis 4 was partially supported.

**Post hoc analyses.** Participants reported 4.63 years of managerial experiences on average, but 33.3% of them had none. Thus, we reestimated our model using only the 120 participants with managerial experience \( \text{mean} = 6.87 \text{ years, SD} = 6.51 \). All findings held with one exception: The effect of derailment potential on protect was no longer significant \( \gamma = -0.24, p = .11 \).

**Discussion**

In two large archival datasets, multisource ratings of managers confirm that although women exhibit slightly fewer ineffective interpersonal behaviors than men do, they are viewed as more likely to derail in the future when poor interpersonal behaviors are observed. This pattern of results, which suggests gender bias, was also found in two experimental studies. Moreover, our final study provides preliminary evidence of the damaging effects that can accrue to managers who are perceived as likely to derail. Not only do perceptions of derailment potential lead to withdrawal of supervisory mentoring, they are most strongly associated with intentions to withdraw sponsorship, which has been closely linked to the upward mobility of women.

Our results clearly and unambiguously refute Landy’s (2008) contention that gender bias is a product of the sterile environment of the laboratory and may not occur in the complexity of the natural work environment where supervisors and employees know each other well. Our data show that evaluations about the future potential of managers are ambiguous enough that shifting standards can creep in, even in a natural work context. In Studies 1 and 2, we found evidence of bias toward women who do not behave consistently with stereotypes prescribing communal behavior. In Study 1, we found gender bias in both directions; there was bias toward women when ineffective interpersonal skills were very low and bias against women when ineffective interpersonal skills were high. However, we were not able to replicate this effect in Study 2.
Landy’s (2008) second argument, that laboratory experiments may inflate the magnitude of bias, was supported by our data. The effects we found in our field studies were very small. On the one hand, our findings reinforce the critical importance of using large archival datasets of working managers and those who evaluate them in their natural setting in resolving this type of debate. On the other hand, the small effects we found were only revealed when examining ambiguous ratings of the future: derailment potential.

*Do Small Effects Matter?*

To some extent, the answer to this question depends on who is asking. From the standpoint of the EEOC, our results are cause for concern, as the same behaviors for men and women are associated with different organizational outcomes. Gender bias, no matter how small, is illegal in the United States and many other countries. From the standpoint of a female manager who is denied the opportunity for sponsorship, small amounts of bias can have severe career consequences. Yet, the small gender × interpersonal behaviors interaction that we found clearly offers only a partial explanation for why women advance more slowly than men do.

Another way to think about the size and importance of our results is to examine the derailment probabilities for men and women who exhibit high levels of ineffective interpersonal behaviors. Let’s assume that a Fortune 500 company with a million people has 500 managers, equally split between men and women, who exhibit high level of ineffective interpersonal behaviors. Based on our tests of practical significant in Study 1, 77% of the women (193), but only 60% of the men (150), are likely to derail. This leaves the organization with 157 managers (of the original 500) in place, but now, rather than an equal number of men and women, we have 100 men and only 57 women remaining. This is a simple illustration, but it has been well established in the literature that even very small effect sizes—indicating tiny but consistent stereotype-based biases in each evaluation—could lead to substantial gender disparities at high levels of an organization (Agars, 2004; Martell, Emrich, & Robison-Cox, 2012), in part because there are fewer gender role models and fewer mentors for subsequent generations of women. Moreover, given the host of personal and contextual variables and the wide range of behaviors that could influence leaders’ perceptions of derailment potential, it is quite impressive that this single set of behaviors had a detectable impact (Cortina & Landis, 2009).

From our research, two things seem clear: (a) Gender bias, based on violation of female stereotypes for communality, can be found in organizations. Moreover, our results suggest that subtle biases in judgments of derailment potential may lead to withdrawal of the type of career
support that is most important to women, and (b) the gender bias effects that we found are too small to be the primary explanation for women’s slow advancement in organizations. Unequal advancement of men and women into upper management positions remains a critical public policy concern, but requires a broader research approach (see Joshi, Neely, Emrich, Griffiths, & George, 2015 for a discussion of this topic).

Contributions and Limitations

In addition to our contribution to the ongoing discussion of stereotype-based gender bias, this research also contributes by taking steps to bridge the science-practitioner gap on the topic of managerial derailment. Derailment of high-potential managers is a key concern for organizations, but scholars have paid only limited attention to the derailment construct, and most research—like our field studies—has used single-item measures. Our research makes an important contribution by beginning the process of construct validation. We provide initial evidence that performance, promotability, and derailment potential represent distinct constructs. We also provide evidence that these constructs are differentially affected by gender bias, as well as preliminary evidence that items used by various leadership consultants represent the same construct. Our Study 4 results suggest that leaders’ perceptions of derailment potential may have important effects on the extent to which managers receive mentoring and sponsorship.

Another strength of this research was the use of multiple methods, which converge on the same pattern of results. Nonetheless, we acknowledge that our findings—though provocative—leave a number of questions open to future research. To begin with, our study incorporated only leaders’ perceptions of the likelihood of career derailment not actual derailment data. Longitudinal field research, in which new managers are followed over time, is critical to developing a better understanding of the antecedents and consequences of leaders’ perceptions of derailment potential. Such research would allow for direct measurement of potential early in a manager’s career, as well as directly observing the effects of male and female managers’ behaviors—good and bad—on changes in perceptions of potential, including derailment potential over time. In addition, clear understanding of derailment potential requires a reliable and well-validated measure of the construct. This is no small task, as the process of derailment unfolds over time and objective measures cannot be used, as it is leaders’ perceptions about the managers’ derailment risks that ultimately influence their actions. That said, it is important to demonstrate some degree of agreement between observers about whether a specific individual is at risk for derailment. In addition, a good measure of derailment potential should include both items that
signal early warning signs of derailment, such as career difficulties or unmet expectations, and items that directly signal more imminent derailment, such as the one used in Study 1. Finally, given the overlap between current job performance, promotability, and derailment potential, qualitative research asking managers directly about these three constructs and what differentiates judgments on each from the others would be useful.

Although our combination of field and experimental studies is a positive feature of the present research, the experimental results represent the reactions of strangers to simulated managerial reports, and are in need of replication in a field sample. Furthermore, our studies do not distinguish whether gender-based expectations directly affect observations of behaviors (“I see something different for a man and a woman”), evaluations of observed behaviors (“I perceive the same behavior to be acceptable for a man but unacceptable for a woman”), or the criteria used for evaluation of derailment potential (“I perceive ineffective interpersonal behaviors to be a stronger signal of derailment for a woman than for a man”). Pinpointing where in the evaluation process standards shift is an important goal for future research.

Our archival field samples also have several limitations. First, although we explored a number of boundary conditions, including the gendered natures of the industry and the job, we did not have data on other potential contextual moderators of these effects. For example, the extent to which ineffective interpersonal behaviors affect derailment perceptions may be influenced by organizational climate for incivility, in addition to gender. Moreover, because managers in our field sample worked for organizations that had both budgets large enough to fund leadership development, and management committed to it, small businesses are not well represented in our data. It may be that the effects we found are less evident—or less damaging—in small organizations, where fewer hierarchical levels and closer relationships between managers at various levels may exist. Finally, many of the managers in Study 1 were recommended for leadership development, which has evolved over time from a way to remediate at-risk managers into a way to support and develop high-potential managers (McCauley & Hezlett, 2002). Thus, our data may not represent the full range of ineffective behaviors or derailment potential that could be found in organizations. Such range restriction, if present, would result in deflation of true effects, suggesting that the small effects we report may be conservative estimates (Raykov & Marcoulides, 2011).

Practical Implications

Some have argued that women avoid seeking sponsorship from senior colleagues because they are reluctant to “use people” and that being
promoted is largely a matter of overcoming this aversion (Hewlett et al., 2010), but our results reveal that women may be unwittingly shut out of opportunities to gain advancement opportunities regardless of their willingness to do so. A natural question is whether managers—particularly female managers—who are perceived as derailment risks can recover from such perceptions and stay on track, or whether perceptions of derailment potential are effectively a “kiss of death” in organizations.

Previous work suggests the former. Practitioner (e.g., Capretta, Clark, & Dai, 2008; Dotlich & Cairo, 2003; Zhang et al., 2013) and research (e.g., Hogan, Hogan, & Kaiser, 2010; Inyang, 2013; Tang, Dai, & De Meuse, 2013; Van Velsor & Leslie, 1995) literatures on derailment converge in arguing that self-awareness, coaching, and leadership development can help managers correct their courses. In case studies, both “Michelle” (Lombardo & Eichinger, 1989) and “Angela” (Capretta, Clark, & Dai, 2008) were female managers regarded as bright but arrogant, alienating peers, or subordinates. With support from organizational mentors who pushed them to identify their ineffective interpersonal behaviors, gave them targeted feedback, and deliberately assigned them to projects that would force them to behave communally to succeed, both women were reported to be back on the fast track.

Individualized personal development is likely to offer benefits for any manager. However, there may be limits to the usefulness of “fix the woman” (de Vries, Webb, & Eveline, 2006) approaches that exhort women to behave more communally to avoid being seen as derailment risks. Female managers are all too aware of the “double bind” that requires them to be simultaneously agentic and not too agentic (Catalyst, 2007). Coaching female managers individually may help them avoid derailing, but it does not address the wider-reaching structural inequities that our study identifies. Furthermore, there is another catch-22 here: case studies suggest that managers who are seen as derailment risks need sponsorship and mentoring to recover, but our findings suggest that leaders may instinctively pull back, not push forward, on these support behaviors.

**Future Research**

We hope that the results of our study will provoke new debate, new research, and new understanding of what holds women back. One important goal for future research is to gain a better understanding of the effects of discrete events on perceptions of derailment potential. These assessments may be shaped by many incidents over time or by single make-or-break failures, depending on the nature of the event and the manager’s position in organizational networks. A related question is what happens to men and women when they make mistakes. There is growing interest in trust repair
in the literature (Tomlinson & Mayer, 2009); future research in this area might examine differences between men and women in how they respond to mistakes and how others respond to them when they make a mistake or break trust.

Another important topic for future research is whether the leadership competencies that lead to promotion and advancement are similar for men and women, because different skills and behaviors may predict derailment and promotion (Hogan et al., 2010). Derailment in particular merits increased research attention, given its practical importance to individuals and organizations. Supervisors’ assessments of derailment potential may have other consequences in addition to the ones we identified. For example, longitudinal field research might assess how supervisors’ withdrawal of mentoring and sponsorship affects subordinates’ networks.

Conclusions

Our goal was to examine the role of shifting standards of evaluations based on gender stereotypes in bosses’ assessments of derailment potential and the impact of those derailment assessments on mentoring and sponsorship behaviors. Our results reveal clear evidence of small but important gender-linked bias in bosses’ assessments of the derailment potential of men and women managers with ineffective interpersonal behaviors. Women who exhibited poor interpersonal behaviors in our samples were evaluated slightly more negatively than men with the same behaviors, especially when both had high levels of ineffective interpersonal behaviors.

Our research highlights the importance of the derailment potential construct for organizations, especially given its potential to elicit gender bias, and the possibility that it may affect mentoring, especially sponsorship behaviors that have subtle but critical implications for female managers’ organizational advancement. Considered as a whole, the results of these studies suggest that gender bias, although subtle, is alive and well—and that it may limit women’s career advancement in more insidious ways than previously understood. Greater understanding of how perceptions of derailment potential are formed and how they influence leaders’ behavior is needed.

REFERENCES


Mumford TV, Campion MA, Morgeson FP. (2007). The leadership skills strataplex: Leadership skill requirements across organizational levels. Leadership Quarterly, 18, 154–166. doi: 10.1016/j.leaqua.2007.01.005


