



**American
Accounting
Association**

**Thought Leaders in
Accounting**

The American Accounting Association is the largest community of accountants in academia. Founded in 1916, we have a rich and reputable history built on leading-edge research and publications. The diversity of our membership creates a fertile environment for collaboration and innovation. Collectively, we shape the future of accounting through teaching, research and a powerful network, ensuring our position as thought leaders in accounting.

Journal of Management Accounting Research
Vol. 34, No. 2
Summer 2022

**It's Complicated: How a Subordinate's Gender
Influences Supervisors' Use of Past
Performance Information When Appraising
Potential**

American Accounting Association | Publications

The American Accounting Association is the copyright holder of this article and retains the worldwide rights to publish, reproduce, distribute, publicly display and/or license the Material in this article for any future editions and versions of this Work, derivative works based on this article (including translations and adaptations), related ancillaries and supplements, and in promotional materials in all print, digital and wireless and/or electronic formats.

Further distribution of this article is strictly prohibited. Written application must be made to the American Accounting Association, 9009 Town Center Parkway, Lakewood Ranch, FL 34202, for permission to reproduce any of the contents of this article, other than for use in courses of instruction. For additional information about the AAA, please visit <http://aaahq.org> where you can also browse the abstracts of AAA journal articles.

It's Complicated: How a Subordinate's Gender Influences Supervisors' Use of Past Performance Information When Appraising Potential

Anne M. Farrell
Michele L. Frank
Miami University

ABSTRACT: Firms often ask supervisors to appraise subordinates' potential to succeed in higher-level positions. These appraisals can take place within social and organizational contexts that emphasize improving the experiences of females in male-dominated settings. Using this context, we experimentally examine whether supervisors differentially interpret and use the same accounting information when appraising the potential of subordinates of different genders. We draw on attribution theory and research suggesting that a diversity emphasis sends two conflicting signals—females are less able than males, but more valuable from a diversity standpoint. We find that supervisors' beliefs about subordinates' abilities in their current positions are lower for female than male subordinates. While prior research suggests this should result in supervisors appraising female subordinates' potential as lower than males', we find the opposite. Our results suggest that firm efforts to improve the workplace experiences of females may do little to mitigate underlying gender stereotypes.

Keywords: performance measurement; performance appraisals; potential appraisals; high-potential employees; gender; attributions.

I. INTRODUCTION

In an effort to identify and retain high-potential employees, upwards of 88 percent of large companies ask supervisors to appraise both subordinates' performance in their current positions and their potential to succeed in higher-level positions (Aon Hewitt 2013; Cappelli and Keller 2014; Chamorro-Premuzic, Adler, and Kaiser 2017; Fernández-Aráoz, Roscoe, and Aramaki 2017; Finkelstein, Costanza, and Goodwin 2018). Appraisals of potential are often used to allocate resources such as leadership training and mentoring (Ibarra, Carter, and Silva 2010; Finkelstein et al. 2018). While managers may consider many factors when appraising potential, 100 percent of the managers surveyed by Church, Rotolo, Ginther, and Levine (2015) say they consider a subordinate's past performance, and 75 percent say it is the most important factor.

Appraisals of potential are not made in a vacuum—they take place within broader contexts. Recently, this context has included societal and organizational pressures to improve the experiences and outcomes of underrepresented groups, including females in male-dominated fields. To address these pressures, at least 85 percent of businesses have explicit diversity initiatives to increase the representation of females in their organizations, especially in higher-level positions (McKinsey 2017; PwC 2019). From an accounting perspective, this raises the question we investigate: Given these societal and organizational pressures, do supervisors consistently interpret and use past performance information in appraisals of potential for subordinates of different genders?

We are grateful for the constructive feedback from Heba Abdel-Rahim, Ted Ahn, Jacob Birnberg, Tina Carpenter, Ivey Che, Mandy Cheng, Willie Choi, Margaret Christ, Ted Christensen, Jackie Hammersley, Dana Hollie, Kerry Humphreys, Theresa Libby, Ella Mae Matsumura, Rich Mautz, Adam Presslee, Jon Pyzoha, Amal Said, Tyler Thomas, Kimberly Walker, Dimitri Yatsenko, James Zhang, and workshop participants at the University of Georgia, UNSW Sydney, The University of Toledo, and the University of Wisconsin. Our thanks also go to Corrine Kidd for her assistance with copyediting, to Michael Macey for his assistance with programming our instrument, and to Kristy Towry for her assistance with the research design and data collection. Anne M. Farrell is grateful for financial support from PricewaterhouseCoopers and the Miami University Department of Accountancy.

Anne M. Farrell and Michele L. Frank, Miami University, Farmer School of Business, Department of Accountancy, Oxford, OH, USA.

Editor's note: Accepted by Andrew H. Newman, under the Senior Editorship of Eva Labro.

Submitted: May 2020
Accepted: December 2021
Published Online: February 2022

We define *potential* as the probability that a subordinate will succeed in a higher-level position (Silzer and Church 2009; Aon Hewitt 2013). We focus on a male-dominated context, meaning workplaces with a higher percentage of male than female employees and that thus tend to be associated with traits stereotypically ascribed to males (e.g., Cejka and Eagly 1999; Heilman 2012). Examples include finance, banking, and STEM fields (Jaekel and St-Onge 2016; Vial and Napier 2018; Braun, Stegmann, Hernandez Bark, Junker, and van Dick 2017; Offermann and Coats 2018; Heilman and Caleo 2018; Zillman 2019). We focus on a male-dominated context because female underrepresentation in this context can translate to significant social and economic disadvantages.¹ Despite diversity initiatives or goals, corporations have made slow progress addressing these disparities—from the entry level to the C-suite, the percentage of females at each level has risen by no more than 4.1 percentage points since 2015, and in 2020 females held just 38 percent of positions at manager levels and above (McKinsey 2020).

Within this context, we examine how subordinate gender impacts supervisors' use of accounting information about past performance when appraising potential. Past performance signals a subordinate's potential to succeed in a higher-level position whenever there is some overlap between the skills required in both positions (Grabner and Moers 2013). Given some overlap, prior accounting research unsurprisingly finds positive relationships between past performance information and supervisors' appraisals of subordinates' potential both to sustain performance in their current positions and to succeed in higher-level positions (e.g., Bol 2008; Bol and Smith 2011; Church and Rotolo 2013; Church et al. 2015; Luft, Shields, and Thomas 2016; Bol and Leiby 2018). Prior research also asserts that the magnitude of the performance–potential relationship depends on supervisors' beliefs about what drove subordinates' past performance. Specifically, the more supervisors attribute past performance to innate abilities (or lack thereof), the higher (lower) the appraisals of potential (Shields, Birnberg, and Frieze 1981; Kaplan and Reckers 1985). Note, though, that this literature is typically silent about how subordinates' gender affects appraisal processes and appraisal outcomes (Johns 2006).

Research in other domains does consider how subordinate gender affects appraisals, particularly, how evaluators rely on categorization based on gender to form beliefs about the drivers of performance (Feldman 1981). In male-dominated contexts, evaluators stereotypically expect males to perform better than females and form beliefs about performance drivers in ways consistent with those expectations (e.g., Deaux and Emswiller 1974; Feldman-Summers and Kiesler 1974; Nieva and Gutek 1980; Heilman 1983; Greenhaus and Parasuraman 1993; Green and Mitchell 1979). When performance is inconsistent with expectations—e.g., when a female's performance is strong or a male's is weak—evaluators tend to attribute a female's strong performance to good luck rather than high ability, and a male's weak performance to bad luck rather than low ability. Thus, in our context, one would expect that because attributions to ability influence appraisals of potential, supervisors would appraise the potential of female subordinates as lower than equally performing males (Greenhaus and Parasuraman 1993).

However, the research above was conducted when society and organizations placed far less emphasis on improving females' workplace experiences and outcomes, so it is valuable to examine how subordinates' gender influences supervisors' appraisals today. We build on a growing body of research to predict that signals about the importance of reducing female underrepresentation in male-dominated contexts can have countervailing effects on how supervisors interpret past performance information and incorporate it into their appraisal ratings (Spence and Keeping 2013; Monnery and Blais 2017; Caleo and Heilman 2019; Leslie 2019; Dover, Kaiser, and Major 2020). On one hand, an emphasis on improving females' workplace experiences and outcomes in male-dominated contexts can unintentionally signal that females need help to succeed. This can lead supervisors to believe that females are less able than their male peers, reinforcing gender stereotypes that shape beliefs about the drivers of performance (Leslie, Mayer, and Kravitz 2014; Caleo and Heilman 2019; Leslie 2019). Thus, we predict that holding past performance constant, the extent to which supervisors attribute strong (weak) past performance to ability (or lack thereof) is lower (higher) for female than male subordinates.

On the other hand, the emphasis on females' experiences and outcomes also signals that diversity is valued and that females are more valuable than males from a diversity standpoint. Because appraisals of potential are a basis for allocating resources designed to retain and develop subordinates, this perception can create upward pressure on supervisors' appraisals of female subordinates' potential (Leslie, Manchester, and Dahm 2017; Krentz 2019; PwC 2019; Lee and Waddell 2021). Thus, holding past performance constant, we predict that despite their underlying beliefs that females are less able than males, supervisors will appraise a female subordinate's potential as being the same as or higher than a male subordinate's. This occurs because supervisors place less weight on their underlying beliefs about subordinates' ability when appraising female versus male subordinates' potential.

¹ This is true because male-dominated contexts are associated with higher status, pay, and upward mobility than female-dominated contexts (Cejka and Eagly 1999; Levanon and Grusky 2016; Hegewisch and Mefferd 2021). In female-dominated contexts like early childhood education, nursing, and social work, there are few social pressures and diversity initiatives to reduce the underrepresentation of males; leadership positions are often held by males even though females dominate lower ranks; and these positions typically come with lower status and pay, so there are fewer social and economic disadvantages to male underrepresentation (Levanon and Grusky 2016; Blau and Kahn 2017; Block, Croft, De Souza, and Schmader 2019).

We test our predictions with a 2×3 between-subjects experiment in which participants with an average of 20 years of professional work experience assumed the role of a wealth management group supervisor. After reading various job descriptions, participants reviewed the past year's performance information and ratings for one subordinate, where we manipulated whether that performance was strong or weak relative to targets. We manipulated the subordinate's gender by either providing no information about gender or information that subtly conveyed the subordinate was male or female. Participants appraised the subordinate's potential to succeed at the next-higher position in the organization, made inferences about the drivers of past performance, and provided other appraisal-related beliefs.

Results were consistent with our predictions. First, when performance was strong, supervisors attributed performance to ability to a lesser extent when the subordinate was female; when performance was weak, supervisors attributed performance to a lack of ability to a greater extent when the subordinate was female. Second, despite viewing females as less able, supervisors appraised a female subordinate's potential as higher, and the difference was significant when past performance was strong. Finally, supervisors placed less weight on their underlying beliefs about the subordinate's ability when appraising a female's potential. Results for the gender-neutral subordinate were the same as those for the male subordinate, confirming that participants stereotypically viewed our context as male-dominated. In a supplemental analysis we consider an alternative explanation for our pattern of results. Taken together, results indicate that supervisors differentially use past performance information when appraising the potential of subordinates of different genders.

Our study makes several contributions to management accounting research. We add to the growing body of research examining how supervisors incorporate past performance measures into appraisals of subordinates' potential in higher-level positions (see, e.g., [Bol and Leiby 2018](#); [Chan 2018](#)). Unlike prior accounting research, we do not hold constant or remain silent about subordinate gender (exceptions are [Lowe, Reckers, and Sanders 2001](#); [Maas and Torres-González 2011](#); [Bloomfield, Rennekamp, Steenhoven, and Stewart 2021](#)). We provide evidence that supervisors may differentially and inconsistently use identical accounting performance information when appraising subordinates of different genders, in contrast to predictions of accounting models that do not consider subordinate genders (e.g., [Shields et al. 1981](#)).

We also provide insights organizations should consider when designing diversity initiatives and appraisal processes. Emphasizing the need to improve the workplace experiences and outcomes of females can lead to more female employees being labeled as high potential, but may do little to eliminate underlying gender stereotypes. This can cause others to believe that females are being rewarded on the basis of their diversity value ([Caleo and Heilman 2019](#)). At best this can lead to differences in resources allocated across male and female high-potential employees, and at worst generate animus toward females and cause females to question their own qualifications and lose confidence and motivation ([Caleo and Heilman 2019](#); [Leslie et al. 2014](#)). Organizations should at least take steps to make biases more salient so supervisors or those higher in rank can try to correct for their effects, and more broadly reconsider the design of initiatives that have done little to mitigate gender stereotypes or increase the representation of females across ranks ([Brescoll, Dawson, and Uhlmann 2010](#); [Ibarra et al. 2010](#); [Dobbin and Kalev 2016](#); [Ryan et al. 2016](#); [Fernández-Aráoz et al. 2017](#); [Heilman and Caleo 2018](#); [McKinsey 2020](#); [Manzi and Heilman 2021](#)).

II. HYPOTHESIS DEVELOPMENT

Background: Why Subordinate Gender May Impact Supervisors' Appraisals of Potential

Prior research recognizes that information about subordinates' past performance is rarely precise, so supervisors use subjectivity, for better or worse, when collecting, interpreting, and aggregating information for appraisals of subordinates' performance in their current positions and their potential to sustain that performance (e.g., see [DeNisi, Cafferty, and Meglino 1984](#); [Prendergast and Topel 1993](#); [Moers 2005](#); [Bol 2008](#); [Bol and Smith 2011](#); [Luft et al. 2016](#); [Bol and Leiby 2018](#); [Demeré, Sedatole, and Woods 2019](#)). Appraisals of subordinates' potential to succeed in different, higher-level positions require even more processing steps, providing latitude for greater subjectivity ([Shah and Oppenheimer 2008](#); [Cappelli and Keller 2014](#); [Adler et al. 2016](#)). First, supervisors should assess the overlap in the abilities required in the current and higher-level positions, but this is challenging since job descriptions or observations of performance rarely provide explicit signals about this overlap ([Mattone and Xavier 2012](#); [Cappelli and Keller 2014](#); [MacRae and Furnham 2014](#); [Benson, Li, and Shue 2019](#)). Next, supervisors should collect and analyze the existing measures of subordinate performance that capture the overlapping abilities. However, there are likely to be some measures of performance that capture abilities that are relevant for the current but not the higher-level position (one subset of non-overlapping abilities); these should be ignored, but ignoring information is notoriously difficult to do ([Nisbett, Zukier, and Lemley 1981](#); [Tetlock, Lerner, and Boettger 1996](#)). Finally, supervisors should collect and analyze information about the extent to which subordinates have the abilities needed for the higher-level position that are not needed for the current position (the other subset of non-overlapping abilities). Information about abilities subordinates do not need and thus are unlikely to regularly exhibit is difficult to capture and measure, if it exists at all ([DeNisi et al. 1984](#); [Fairburn](#)

and Malcomson 2001; Mattone and Xavier 2012; Church and Rotolo 2013; Cappelli and Keller 2014; MacRae and Furnham 2014).

Models of appraisal processes suggest that when relevant information is more difficult to identify and analyze, highly salient information, even if irrelevant, becomes more important to judgment processes and outcomes (Ilgen and Feldman 1983; DeNisi et al. 1984; Ilgen, Barnes-Farrell, and McKellin 1993). In a context where females are underrepresented, we expect subordinates' gender will be particularly salient and thus likely to influence appraisal processes and outcomes (Heilman 1983; Ilgen and Feldman 1983; Zarate and Smith 1990).

Appraisal Processes

Because subordinates' performance information rarely provides precise signals of their contributions to firm success, supervisors must subjectively discern the drivers of a subordinate's performance (Birnberg, Frieze, and Shields 1977; Shields et al. 1981; Banker and Datar 1989; Prendergast and Topel 1993; Feltham and Xie 1994; Lambert 2001). Supervisors then rely on these beliefs (and any other information they identify as relevant) as an input into appraisal ratings formally reported to the subordinate and others in the organization.

Prior attribution theory research finds that supervisors typically attribute subordinates' performance to four factors—ability, effort, good or bad luck, and job difficulty. These attributions differ in terms of their locus of control (i.e., internal or external to the subordinate) and the likelihood that they will fluctuate over time. Ability and effort relate to a subordinate's internal, dispositional, or behavioral characteristics, while luck and job difficulty relate to external, situational factors. Ability and job difficulty are viewed as being relatively stable over time, while effort and luck are viewed as highly variable and temporary (Heider 1958; Weiner and Kukla 1970; Weiner et al. 1971; Kelley 1973; Birnberg et al. 1977; Mitchell and Wood 1980; Shields et al. 1981; DeNisi et al. 1984; Harrison, West, and Reneau 1988).

Prior research finds that gender stereotypes influence performance attributions (Deaux and Emswiller 1974; Green and Mitchell 1979; Heilman 1983; Biernat and Fuegen 2001; Bigelow, Lundmark, McLean Parks, and Wuebker 2014). Evaluators develop stereotypic expectations about how well a task will be performed by a member of a particular group (e.g., females or males), and attribute performance in ways that allow them to maintain their stereotypic beliefs. Specifically, when a person's performance is consistent with the evaluator's stereotypic expectations, the evaluator believes it was driven by an internal, stable characteristic (e.g., confirming expectations about the person's ability or lack thereof). However, when performance is inconsistent with stereotypic expectations, evaluators find it difficult to believe performance could have been driven by a stable internal cause, and instead assume performance was driven by a temporary external cause beyond the person's control. Consistent with this pattern, a significant stream of research finds that females' successes in male-dominated contexts tend to be attributed to good luck and males' successes to high ability. Conversely, males' failures tend to be attributed to bad luck and females' failures to a lack of ability (see, e.g., Deaux and Emswiller 1974; Green and Mitchell 1979; Heilman 1983; Brewer 1988; Greenhaus and Parasuraman 1993; Foschi 2000; Lyness and Heilman 2006).

Much of this research was conducted when society and organizations placed less emphasis on improving the workplace experiences and outcomes of females in male-dominated contexts. In recent years, society and—by extension, organizations—have become increasingly focused on reducing underrepresentation of members of a variety of groups, including females (Society for Human Resource Management 2009; McKinsey 2017, 2020; Brosnan 2018; Scarborough 2018; Shook and Sweet 2018; Krentz 2019; PwC 2019; Zillman 2019; Fuhrmans 2020). For example, a 2017 McKinsey study indicates that 85 percent of companies track gender representation across levels of their organizations, and 90 percent view gender diversity as a high priority. A 2019 PwC survey notes that 87 percent of respondents' companies have greater diversity of underrepresented groups as an explicit value or priority, and prescribes that “holding leaders accountable for D&I [diversity and inclusion] results” and “embedding a diversity lens into talent management” are essential for progress on diversity and inclusion.

While it is possible that today's context has lessened the impact of underlying gender stereotypes on attributions of the drivers of past performance, recent research suggests this may not be true. Specifically, the current societal and organizational focus on improving females' workplace experiences and outcomes may unintentionally signal that females need help to succeed in male-dominated contexts, reinforcing a gender stereotype that females are less able than equally performing males (Tinkler 2013; Dobbin and Kalev 2016; Leslie 2019). Any signal reinforcing the application of underlying stereotypes is unlikely to do much, if anything, to mitigate this application when drawing inferences about the drivers of performance. Further, a significant stream of research finds that males and females hold the same gender stereotypes, including that males are more able than females in male-dominated contexts, so the supervisors' own gender is not likely to influence their beliefs about the drivers of subordinates' performance (see Heilman 2012 for a review). Given this, we expect to replicate prior research that subordinates' gender will influence how supervisors interpret past performance information when forming beliefs about the drivers of past performance.

H1a: In a male-dominated context, supervisors' attributions of strong past performance to ability will be lower when the subordinate is female rather than male.

H1b: In a male-dominated context, supervisors' attributions of weak past performance to a lack of ability will be higher when the subordinate is female rather than male.

Appraisal Outcomes

Recall that past performance is informative about a subordinate's ability to succeed in a different, higher-level position whenever overlapping skills are required in the current and higher-level positions (Grabner and Moers, 2013).² In such cases, past performance should be positively associated with a supervisor's appraisal of the subordinate's potential to succeed in the higher-level position. Further, prior research finds that beliefs about drivers of subordinates' past performance predictably influence the strength of this relationship. That is, supervisors' appraisals of subordinates' potential should be higher (lower) the more supervisors attribute subordinates' past successes (failures) to ability because internal, stable drivers of performance are more likely to be predictive of future successes (failures) than those that are situational or highly variable (Weiner et al. 1971; Valle and Frieze 1976; Birnberg et al. 1977; Heilman and Guzzo 1978; Green and Mitchell 1979; Shields et al. 1981; Pence, Pendleton, Dobbins, and Sgro 1982; Kaplan and Reckers 1985; Pazy 1987; Harvey, Madison, Martinko, T. R. Crook, and T. A. Crook 2014). Thus, one might expect that given our first hypothesis, appraisals of potential would be lower for females than males.

Also recall that prior accounting research on appraisals largely ignores signals triggered by the broader context of the appraisals and thus does not consider the effects these signals can have on supervisors' ratings. In addition to the signal it sends about female subordinates' abilities, a societal and organizational focus on fostering better experiences and outcomes for females also signals that reducing the underrepresentation of females is valued by society and organizations (Ostroff and Bowen 2000; Bowen and Ostroff 2004; Leslie et al. 2017; Leslie 2019). This causes supervisors to consider subordinates' differential value from a diversity standpoint. Specifically, the more likely it is that a particular subordinate's presence in a higher-level position will create diversity, the more valuable that subordinate is from a diversity perspective. Indeed, in a hiring setting, Lee and Waddell (2021) find that organizations' diversity preferences can influence the decisions of supervisors.

Furthermore, prior research finds that supervisors value subordinates in this way regardless of their own personal beliefs regarding the value, or lack thereof, of diversity. That is, if a particular subordinates' presence creates diversity, individuals tend to view them as valuable from a diversity perspective regardless of their own personal beliefs about the value of a diverse workplace (Leslie et al. 2017). Therefore, even though females value diversity more than males (Pew Research Center 2019), we expect that in male-dominated contexts both male and female supervisors will view female subordinates as more valuable from a diversity perspective than equally performing male subordinates (Leslie 2019).

Studies suggest that because females are viewed as more valuable than males from a diversity perspective and appraisals of potential are a basis for allocating resources designed to retain and develop subordinates, knowing that one is appraising a female subordinate can create upward pressure on appraisals of potential (Leslie et al. 2017). Supervisors' underlying reasons for acting on that pressure can vary. For example, supervisors may indeed share and want to promote their organizations' diversity norms and goals, or at least appear as if they do even if they consciously believe females are less able than males; they may fear backlash or reprisal if they do not rate females equal to or higher than males; or the societal and organizational emphasis on diversity may drive them to consider and perhaps overweight or counteract the effect of what they believe to be their own gender biases. Regardless of the reason, the directional effect is the same. Therefore, we predict supervisors' appraisals of female subordinates' potential will be more favorable than predicted by prior accounting models, given supervisors' underlying stereotypical beliefs that females' ability is lower than males'. This upward pressure is likely strongest when a female subordinate's performance is strong, because it is difficult to justify labeling any weak-performing subordinate, regardless of diversity value, as having high potential (Leslie et al. 2017). Based on this, we predict that despite viewing females as less able, supervisors will appraise a female subordinate's potential to be the same as or higher than a male subordinate's.

H2a: In a male-dominated context, supervisors will appraise a female subordinate's potential to succeed in a higher-level position as being the same as or higher than an identically performing male subordinate's potential.

² While supervisors likely also incorporate their beliefs about subordinates' abilities with respect to skills needed for the higher-level position that fall outside any skills overlap, literature on the use of subjectivity in appraisals finds that subjective assessments about such abilities follow the same patterns that we predict with respect to beliefs about subordinates' abilities needed for their current positions (Koenig, Eagly, Mitchell, and Ristikari 2011; Schein and Davidson 1993; Vial and Napier 2018).

We predict that these differential appraisals of potential for identically performing subordinates occur because supervisors place less weight on their underlying beliefs about subordinates' ability when appraising female versus male subordinates. In other words, supervisors' desire to comply with societal and organizational values suppresses the influence of their stereotypical beliefs about females' ability when providing formal appraisals of potential.

H2b: Supervisors' beliefs about a subordinate's ability will have less of an influence on the supervisor's appraisals of the subordinate's potential when the subordinate is female rather than male.

III. METHOD

Participants

We used business school alumni to proxy for the population of business professionals experienced in using performance information to appraise subordinates and likely to be at least minimally aware of the current societal and organizational focus on improving the experiences and outcomes of females working in male-dominated fields. We recruited alumni to participate in the study with the assistance of a business school's alumni relations office. Alumni relations staff retained control of email addresses, delivery schedules, and content and approved the experimental materials to ensure that alumni were not overwhelmed with solicitations and that participants remained anonymous. Participants did not receive compensation.

Alumni staff emailed a random sample of about 4,500 alumni, asking those with two or more years' experience managing employees to participate in a research study; a follow-up email was sent about one week later.³ The emails included a link to a Qualtrics survey that randomly assigned participants to one of seven experimental conditions (described later) when they clicked on the link. Of 301 individuals who clicked on the link, 172 completed the study. Of those, 160 were representative of the population of interest and completed the task (a response rate of about 3.6 percent of emails sent and 53.2 percent of individuals who clicked on the link).^{4,5}

Consistent with our assumption that this sample was a sound proxy for the population of interest, the 160 participants have an average of 20.3 years of professional work experience, and approximately 87.5 percent list job titles that indicate a supervisory role (CEO, CFO, director, president, vice president, partner/principal, manager). Participants' average age is 44.0 years, and 25.6 percent are female (Table 1). While we would have liked to have a higher percentage of female participants, our percentage is consistent with the percentage of women holding leadership positions in practice. While women make up nearly half of the U.S. labor force, they only hold about 24 percent of senior roles (e.g., manager, director, vice president, president; see, e.g., [Brosnan 2018](#); [Catalyst 2018](#)).

Experimental Setting and Procedures

We used a hypothetical case set in the banking and financial services industry to test our hypotheses.⁶ We chose this setting because both males and females view this industry, and most of the jobs within it, as male dominated ([Alden 2014](#); [Zillman 2019](#)). This is unsurprising, given that a majority of professionals within the industry are male. Females make up 32.1 percent of wealth management employees across all levels, with just 20 percent of higher-level roles held by women ([World Economic Forum 2016, 2017](#); [Catalyst 2018](#)). In addition, both men and women tend to associate professionals in the industry with attributes (e.g., achievement oriented, rational, analytical) that are stereotypically associated with males ([Alden 2014](#); [Bigelow et al. 2014](#); [Jaekel and St-Onge 2016](#); [World Economic Forum 2016, 2017](#); [Catalyst 2018](#); [Heilman and Caleo 2018](#); [Holman, Keller, and Colby 2018](#); [Zillman 2019](#)).

³ We collected data in July 2017, when diversity programs were commonplace but before October 2017 when the #MeToo Movement began its viral spread and made workplace gender equity even more salient ([Johnson and Hawbaker 2019](#)); this biases against H2a and H2b relative to collecting data after that point.

⁴ We compare these rates to those in studies that sought participants with significant work experience, but note that because our solicitation was handled entirely by the alumni association, we do not have access to bounce or open rates of the emails (as other studies do). The percentage of usable responses relative to emails sent, inclusive of bounces, was 0.7 percent in [Anderson and Lillis \(2011\)](#); more than 10,000 emails sent) and 1.6 percent in [Dichev, Graham, Harvey, and Rajgopal \(2013\)](#); more than 10,000 emails sent); the percentage was 7.7 percent in [Graham, Harvey, and Puri \(2013\)](#); more than 29,000 emails sent), but the reported number of emails sent excluded bounces. [Anderson and Lillis \(2011\)](#) report the percentage of usable responses out of links opened was 72.17 percent.

⁵ We eliminated 12 participants' responses since they reported work experience or job titles that were not representative of the population of interest (e.g., summer intern or entry-level consultant; "mother, wife, chef, CEO of our home"). One participant did not complete the task.

⁶ The study received Institutional Review Board approval. Our case is based on the private wealth management group of a global banking and financial services company headquartered in New York City. Two of the group's Senior Vice Presidents reviewed our case materials to ensure they accurately reflected the way in which potential is judged within their organization. In addition, the position descriptions are consistent with descriptions in a survey of the skills required in portfolio manager and team leader roles ([Sales Management Association 2008](#)).

TABLE 1
Participant Demographic Data and Task-Related Beliefs

Panel A: Participant Demographic Data

	<u>Mean</u>	<u>Std. Dev.</u>	<u>Count</u>	<u>Percent</u>
Age, in years	44.03	9.54		
Professional work experience, in years	20.31	8.80		
Participant Gender				
Male			119	74.40%
Female			41	25.60%
Total			160	100.00%
Job Title				
CEO			9	5.62%
CFO			7	4.37%
Director			54	33.75%
President			3	1.87%
Vice President			25	15.63%
Partner/Principal			13	8.13%
Manager			29	18.13%
Other			20	12.50%
Total			160	100.00%

Panel B: Mean (Standard Deviation) of *Overlap* by Condition

	<i>Strong Past Performance</i> ^b			<i>Weak Past Performance</i> ^b			<i>Control</i> n = 25
	<i>Gender</i> ^c			<i>Gender</i> ^c			
	<i>Male</i>	<i>Female</i>	<i>Neutral</i>	<i>Male</i>	<i>Female</i>	<i>Neutral</i>	
	n = 22	n = 23	n = 21	n = 25	n = 21	n = 23	
<i>Overlap</i> ^a	4.32 (1.32)	4.04 (0.93)	4.10 (1.04)	3.40 (0.91)	3.33 (0.91)	3.61 (1.20)	4.08 (1.26)

^a We measured participants' beliefs about the extent of overlap in the skills required to be an effective portfolio manager and team leader with this question, anchored on 1 = No overlap and 7 = Complete overlap, with a midpoint 4 = Some overlap: "To what extent do the skills required to be an effective portfolio manager overlap with the skills necessary to be an effective team leader?"

^b We manipulated *Past Performance* by providing actual performance and targets for five quantitative performance measures, above- or below-par ratings for each measure, performance notes, and an overall above- or below-par performance rating. *Strong (Weak) Past Performance* was operationalized as a majority of above-par (below-par) ratings for the five quantitative performance measures and an overall above-par (below-par) rating.

^c We manipulated *Gender* by including male, female, or gender-neutral wording and silhouettes just before the subordinate's past performance information, which was either *Strong* or *Weak*. In *Control*, participants reviewed the same background information as participants in the *Gender* conditions, but only answered the *Overlap* question; they did not receive any subordinate performance information or make appraisals of potential or causal attributions.

After providing informed consent, participants were given an overview of a bank's private wealth management group and the hierarchy of positions within the group (Appendix A), including portfolio manager and team leader positions. Participants were told that portfolio managers "work directly with clients, helping them manage their investments, income, and estate planning" and "are responsible for attracting new clients." Team leaders "manage teams of 7–10 portfolio managers" and "are responsible for managing resources within their team and making sure that the portfolio managers who report to them understand and act in ways consistent with the bank's strategy."

Two features of the position descriptions are important. First, both describe responsibilities that require communal skills that are stereotypically associated with females (e.g., working with current and attracting new clients; managing subordinates) and analytic skills that are stereotypically associated with males (e.g., managing clients' wealth and estates; managing team resources). Theoretically, past performance provides a signal about potential when there is at least some overlap in the skills required for the two positions under consideration (Grabner and Moers 2013). By including responsibilities that require communal skills and analytic skills in both position descriptions, participants were more likely to perceive there was at least

some overlap in the skills required for the two positions. We chose not to explicitly convey information about skills overlap to participants since such information is rarely, if ever, provided to evaluators in practice. In addition, including job responsibilities that require both communal and analytic skills in both position descriptions biases against participants stereotypically believing that females are better suited for the Team Leader position. Second, we intentionally used the word *Leader* in the title for the higher-level position since both men and women stereotypically associate leadership roles, especially those in male-dominated industries, with males, biasing against H2a (Heilman, Block, Martell, and Simon 1989; Schein and Davidson 1993; Schein 2007; Johnson, Murphy, Zewdie, and Reichard 2008; Koenig et al. 2011; Badura, Grijalva, Newman, Yan, and Jeon 2018; Offermann and Coats 2018; Fuhrmans 2020).⁷

Participants in our treatment conditions are told to assume that they are a team leader. They were then told that the bank experienced an unexpected economic decline in the last three quarters of the prior year. We included this wording so it would not be surprising that subordinates' past performance was worse than targets, biasing toward attributions to ability for subordinates who performed well and against attributions to a lack of ability for subordinates who performed poorly. Participants were then told that one of their responsibilities was to evaluate the performance of the portfolio managers who report to them and were given the past year's performance information and ratings for one of their portfolio managers for the purpose of appraising that portfolio manager's potential to succeed as a team leader.

The performance information included five quantitative performance measures that participants without banking or wealth management experience could understand (e.g., number of new clients), along with the subordinate's actual and target performance and performance rating (*below*, *at*, or *above par*) for each measure and an overall performance rating. We provided ratings associated with the measures rather than asking participants to provide their own ratings to bias against finding different attributions of the drivers of performance for subordinates of different genders. This information also included three qualitative performance notes that provided mixed signals regarding the subordinate's performance. Specifically, these notes stated that the subordinate takes time to adapt to change, communicates in a direct and straightforward manner, and seeks advice when appropriate. While we manipulated whether the five quantitative performance measures and associated ratings and the overall performance rating indicated the subordinate was a strong or weak performer, as described below, the content of the three performance notes was held constant across conditions.⁸

After reviewing the performance information, participants appraised the portfolio manager's potential to succeed as a team leader. Participants also responded to questions that captured their beliefs about the overlap in the skills required for the portfolio manager and team leader positions, the relevance of past performance information for appraisals of potential, and the drivers of the portfolio manager's past performance. Finally, participants responded to demographic questions about their gender, age, education, and work experience.

We intentionally chose not to explicitly ask participants about the value of diversity or diversity initiatives, whether they place a value on subordinates from a diversity standpoint, or whether they have felt pressure to appraise female and male subordinates differently. First, responses to such questions would undoubtedly be biased; it is unlikely that respondents would rate the value of diversity low or agree that they do not value diversity or have felt pressure to rate female and male subordinates differently. This is especially true with our population. While responses were anonymous to researchers, participants were solicited by their university alumni association, so those who bothered to open and read the email likely had strong social ties with the institution. Thus, we expected that responses would be clustered at socially desirable ends of the ratings scales with little variance, providing little insight into participants' actual beliefs. Second, prior research suggests that thorny questions can make participants uncomfortable or create pressure to respond in ways that distort their true beliefs (e.g., Nederhof 1985; Paulhus and Reid 1991; Tourangeau and Yan 2007). In deference to alumni relations staff who are keen to avoid questions that could impact attitudes toward the business school or responses to their later requests, we believed the value of obtaining judgments from experienced professional participants outweighed the potential cost of negatively impacting future alumni engagement or collecting biased responses to questions about the value of diversity or pressures to upwardly-rate females.

⁷ While communal traits such as concern for others and interpersonal sensitivity, which are stereotypically believed to be held by females more than males, may actually be crucial leadership attributes, both males and females tend to view communality as "a non-vital complement to the fundamentally masculine core of the leader role" (Vial and Napier 2018, 9).

⁸ We chose to provide measures of past performance and performance notes rather than or in addition to forward-looking information about skills required for the higher-level but not the current positions. First, upwards of 75 percent of companies use appraisals of past performance to identify high-potential employees (Church and Rotolo 2013; Church et al. 2015). Second, many organizations do not explicitly define competencies for specific positions, nor do they measure subordinates' competencies for higher-level positions prior to promoting them (Church et al. 2015; Benson et al. 2019). Third, even when explicit descriptions of a subordinate's abilities unique to the higher-level position are available, supervisors still focus on quantitative measures of past performance when appraising potential to avoid perceptions of inconsistency, influence activities, and favoritism (Chan 2018; Prendergast and Topel 1993).

Independent Variables

We conduct an experiment using a $(2 \times 3) + 1$ between-subjects design. We manipulated subordinates' *Past Performance* at two levels—*Strong* and *Weak*. *Strong* (*Weak*) performance was operationalized as a majority of above-par (below-par) ratings for the five quantitative performance measures and an overall above-par (below-par) rating (Appendix B). To create as little variation as possible in the past performance measures and ratings seen by participants, we created *Weak* from *Strong* by flipping the numeric values of actual performance and of targets so both the magnitude of the amounts and the percentage differences between actual and target were the same, and the measure-by-measure and overall performance ratings were mirror images. All textual performance notes were identical across *Strong* and *Weak*.

We manipulate subordinate *Gender* at three levels—*Male*, *Female*, and *Neutral*. In *Male* (*Female*), we included the name Thomas (Jennifer) Roan in a sentence and a small, black-and-white male (female) silhouette just above the past performance information. In *Neutral*, we used a sentence without a name and a gender-neutral silhouette (Appendix B); we use these conditions to test whether participants spontaneously viewed our setting as male-dominated.^{9,10}

We consciously chose not to manipulate the presence of diversity initiatives or goals. We concluded it would be impossible to describe a diversity program or state that the hypothetical bank emphasized improving females' workplace outcomes without making it obvious to participants that we were interested in their reactions to subordinate gender, biasing responses in a socially desirable direction and creating demand effects (e.g., Nederhof 1985; Paulhus and Reid 1991; Tourangeau and Yan 2007). As such, conditions in which it was stated that there was a diversity initiative or emphasis could not be used to cleanly test whether patterns of past performance attributions identified decades ago replicate today (not to mention that those studies did not include such statements), as predicted in H1a and H1b. Importantly, this choice biases against finding results for H2a and H2b within *Strong* and *Weak* in our male-dominated context. We acknowledge these choices create the possibility of an alternative stereotype-based explanation for our predicted pattern of results; we address this in supplemental analysis.

Dependent Measures

Participants in our six treatment conditions were asked to appraise the portfolio manager's potential and to provide their beliefs about what drove the portfolio manager's past performance. We measured the appraisal of potential (*Potential*) with this question, anchored on 1 = Strongly disagree and 7 = Strongly agree, with a midpoint 4 = Neutral: "To what extent do you agree that [Thomas Roan/Jennifer Roan/the portfolio manager] has what it takes to succeed if [he/she/he or she] is promoted to the next highest level within the bank?"

We captured participants' causal attributions of the drivers of the subordinate's past performance by posing this question: "To what extent did the following factors contribute to [Thomas Roan's/Jennifer Roan's/the portfolio manager's] above [or below] par performance?" Participants allocated 100 points across five factors: skill/ability (*Ability*), effort (*Effort*), luck (*Luck*), performance target not at appropriate level (*Unreasonable Target*), and other (*Other*, with an open-ended response box). Participants were instructed to allocate more (less) points to factors they believed were more (less) significant drivers of past performance.

Participants in all conditions provided their beliefs about the overlap in the skills needed in the portfolio manager and team leader positions (*Overlap*) by responding to the following question using a scale anchored on 1 = No overlap and 7 = Complete overlap, with a midpoint of 4 = Some overlap: "To what extent do the skills required to be an effective Portfolio Manager overlap with the skills necessary to be an effective Team Leader?"¹¹

⁹ We intentionally chose a between-subjects rather than within-subjects manipulation for *Gender*. A within-subjects manipulation that included only a few subordinates (in the interest of our experienced professional participants' time) would make it obvious to participants that we were interested in reactions to subordinate gender, thus biasing responses in a socially desirable direction. A between-subjects design allowed us to include subordinate gender information in a subtler way among other task information and questions. Also note that asking participants to allocate resources based on their appraisals of potential (e.g., when allocating a raise pool or deciding which subordinate(s) to promote) would necessarily require a comparison across subordinates and a within-subjects design to simulate constraints in the resources that could be allocated to any one subordinate.

¹⁰ We also collected data for a *Control* condition. These participants read the same descriptions of the wealth management group and hierarchy of positions as did participants in treatment conditions, but they did not receive past performance information or complete an appraisal for a portfolio manager. *Control* participants responded to the same questions about the positions and demographics as did participants in our six treatment conditions. Thus, *Control* provides a gender-neutral, appraisal-free baseline for beliefs about aspects of the appraisal process.

¹¹ Participants also responded to the question, "To what extent do you believe that [Thomas Roan's/Jennifer Roan's/a portfolio manager's] past performance is the best predictor of how well [he/she/he or she] will perform as a team leader?" using a scale anchored on 1 (7) = *Strongly disagree* (*agree*). Not surprisingly, responses were significantly correlated with *Overlap* ($p < 0.01$) since the extent to which past performance is predictive of potential depends on beliefs about skills overlap. Using responses to this question instead of *Overlap* in analyses in Section IV results in the same statistical inferences. Thus, we do not use this measure in analyses.

IV. RESULTS

Preliminary Analyses

We expect the difference in attributions to ability for female and male subordinates to change sign across *Strong* and *Weak Past Performance* conditions—attributions to ability will be lower for females than males in *Strong*, while attributions to (lack of) ability will be higher for females than males in *Weak*. Thus, for ease of explication, we present results separately for *Strong* and *Weak Past Performance* conditions.¹² All reported p-values are two-tailed.

Effects of Participant Demographic Differences

For all statistical tests that follow, inferences are the same when we control for participants' age. The same is true when we control for participants' gender—as expected, and consistent with prior research that both males and females hold the same gender stereotypes (Heilman 2012; Hentschel, Heilman, and Peus 2019).¹³ This suggests both males and females are susceptible to differentially interpreting and using past performance information when appraising subordinates of different genders. However, given that each condition has fewer than 10 female participants (the total across conditions is consistent with female representation in our population of interest, as reported in Section III), failure to find an effect of gender may be due to lack of power. Regardless, within each *Past Performance* × *Gender* condition, we compared past performance attributions, *Potential*, *Overlap*, and *Predictive Power* across male and female participants. The only significant differences ($p < 0.05$) were within *Weak*, where *Unreasonable Target* differed across male and female participants in *Female*, and *Overlap* differed across male and female participants in *Neutral*.

Beliefs About Skills Overlap for Two Positions

As noted previously, accounting information about past performance is a useful signal for appraising potential when there is at least some overlap in the skills required for the current and higher-level positions (Grabner and Moers 2013). Those skills are dependent on each job's responsibilities, not on subordinates themselves. Thus, operationally, we described the responsibilities of the current and higher-level positions in a way that we hoped would prompt participants to perceive at least some overlap in the skills needed to be successful. We use *Overlap* to confirm whether participants did so and that those beliefs did not differ across *Gender* conditions. Descriptive statistics for *Overlap* in all conditions are in Table 1, Panel B.

Across all conditions, only three of 160 participants (1.9 percent, untabulated) believed there was no overlap in the skills required for the portfolio manager and team leader positions (i.e., responses of 1 = No overlap), and no participants believed there was complete overlap (responses of 7 = Complete overlap). While remaining responses tend toward the middle of the scale (i.e., 39.4 percent of all responses are 4 = Some overlap), there is a reasonable distribution of responses on both sides of the midpoint (i.e., ratings of 2 and 3 comprise 33.1 percent of all responses and ratings of 5 and 6 comprise 25.6 percent of all responses; untabulated). The distribution of responses does not significantly differ across the three *Gender* conditions (Kruskal-Wallis test, $p > 0.35$ in both *Strong* and *Weak*).

We confirm that beliefs about the skills overlap were not impacted by subordinate gender. The distributions of *Overlap* responses does not significantly differ across the three *Gender* conditions in *Strong* or *Weak* (independent samples Kruskal-Wallis test, both $p > 0.35$, untabulated). Also, results of ANOVAs with *Overlap* as the dependent variable and *Gender* as the independent variable are not significant in *Strong* or *Weak* ($F = 0.38$ and 0.66 , respectively, both $p > 0.50$; untabulated).¹⁴ Overall, this indicates that the majority of participants believed there was at least some overlap in the skills required for the two positions, and these beliefs were not impacted by the subordinate's gender.

¹² Despite the complexity these sign differences introduce when interpreting 2×3 ANOVAs, we report them for completeness. In a 2×3 ANOVA with *Ability* as the dependent variable, *Past Performance* and the *Past Performance* × *Gender* interaction were significant (both $p < 0.01$); when *Potential* was the dependent variable, *Past Performance*, *Gender*, and the *Past Performance* × *Gender* interaction were significant (all $p < 0.10$). All assumptions were met for all ANOVAs.

¹³ Heilman (2012) notes that the fact that male and female evaluators hold the same stereotypical beliefs is puzzling, but suggests that the reasons underlying those beliefs may differ. For example, in male-dominated contexts, males may have a vested interest in maintaining the *status quo*, while females may respond negatively to other females, especially successful ones, because of social comparisons. Further investigation of this puzzle is beyond our scope.

¹⁴ In our *Control* condition, in which beliefs are gender-, performance-, and appraisal-free, *Overlap* did not differ from the *Strong* conditions, but was higher than in the *Weak* conditions (Table 1, Panel B). We expect this is due to supervisors making less effort to assess skills overlap for weak-performing subordinates who are unlikely to be viewed as having high potential. Regardless, we include *Overlap* in all analyses to control for the relevance of these beliefs to appraisals of potential.

Participants' Perception of Male-Dominated Context

We confirm whether participants viewed our context as male-dominated. If they did, participants in *Neutral* should make judgments consistent with those made by participants in *Male*. Thus, we compare attributions to ability (H1a and H1b), judgments of potential (H2a), and the relationship between the two (H2b) in *Neutral* conditions to those made in the corresponding *Male* conditions. While details of these tests are in footnotes with tests of hypotheses and in tables, we note here that there were no differences in judgments across *Male* and *Neutral*, but there were differences between those conditions and *Female*. Thus, it appears that participants spontaneously viewed our setting as male-dominated and/or assumed the subordinate was male when no information about gender was provided.

H1a and H1b

H1a predicts that the extent to which supervisors attribute strong past performance to ability will be lower when the subordinate is female versus male; H1b predicts that the extent to which supervisors attribute weak past performance to lack of ability will be higher when the subordinate is female versus male. Table 2, Panel A, presents descriptive statistics for attributions of past performance in the *Strong* and *Weak* conditions. We tested each hypothesis with an ANOVA and planned comparisons of *Ability*.

For H1a, ANOVA results for *Strong* (Table 2, Panel B) showed a significant effect of *Gender* on *Ability* ($F_{2,65} = 5.52$, $p = 0.01$). Planned comparisons (Table 2, Panel C) showed that attributions to *Ability* were significantly lower in *Female* (mean = 36.35) than *Male* (mean = 48.41; $t = -2.70$, $p = 0.01$). When we compared attribution of performance to factors other than *Ability* across *Male* and *Female*, the only significant difference was *Luck*, with a mean of 7.73 points in *Male* and 15.83 points in *Female* (Table 2, Panel A; $p = 0.06$, not tabulated). Thus, consistent with H1a and prior literature, our participants attribute strong performance in ways consistent with stereotypic expectations. That is, they are more likely to attribute a male's versus a female's strong performance to ability (an internal, stable factor) and a female's versus a male's strong performance to good luck (an external, unstable factor).

For H1b, ANOVA results for *Weak* (Table 2, Panel D) also showed a significant effect of *Gender* on *Ability* ($F_{2,68} = 7.60$, $p = 0.01$), but in the opposite direction of that observed in *Strong*, which represents attributions to a lack of ability. Planned comparisons (Table 2, Panel E) show that attributions of performance to a lack of ability were higher in *Female* (mean = 42.05) than *Male* (mean = 25.80; $t = 3.33$, $p = 0.01$). When we compared attributions to other drivers of performance across *Male* and *Female*, the only significant difference was in *Unreasonable Target*, with a mean of 31.20 points in *Male* and 10.71 points in *Female* (Table 2, Panel A; $t = 3.04$, $p = 0.004$, not tabulated). As was the case when performance was strong, participants attribute weak performance in ways consistent with stereotypic expectations. Specifically, they attributed males' weak performance to unreasonable targets (external factor beyond the subordinate's control) and females' weak performance to lack of ability (internal, stable factor); the latter is consistent with H1b.¹⁵

Overall, results support H1a and H1b. This suggests that despite today's societal and organizational focus on improving the workplace experiences and outcomes of females, underlying gender stereotypes persist in appraisal processes, and thus the same past performance information is not interpreted consistently across subordinates of different genders.

H2a

H2a predicts that supervisors will judge a female subordinate's potential to succeed in a higher-level position as being the same as or higher than an identically performing male subordinate's potential. Table 3, Panel A, presents descriptive statistics for *Potential*.

We tested H2a with an ANOVA and planned comparisons of *Potential*. Table 3, Panel B, reports the ANOVA for *Strong*, which finds a significant effect of *Gender* on *Potential* ($F_{2,65} = 3.15$, $p = 0.05$). Planned comparisons (Table 3, Panel C) showed that *Potential* was significantly higher in *Female* (mean = 5.35) than *Male* (mean = 4.41; $t = 2.65$, $p = 0.01$).¹⁶ Thus, despite viewing females as less able than males, participants appraised a female subordinate's potential as being higher than an equally-strong-performing male.

¹⁵ We compare attributions in *Neutral* to those in corresponding *Male* conditions. The second planned comparisons in Table 2, Panels C and E, show that *Ability* did not differ in *Male* and *Neutral* (*Strong* $t = -0.23$, $p = 0.82$; *Weak* $t = -0.40$, $p = 0.70$). *Effort*, *Luck*, *Unreasonable Target*, and *Other* also did not differ in *Male* and *Neutral* in *Strong* or *Weak* (all $p > 0.56$, untabulated). However, consistent with our comparisons of *Male* and *Female*, the third comparison in Table 2, Panel C, finds that in *Strong*, *Ability* was significantly lower in *Female* (mean = 36.35) than *Neutral* (mean = 49.52; $t = -3.28$, $p < 0.01$), and the third comparison in Table 2, Panel E, finds that in *Weak*, (lack of) *Ability* was significantly higher in *Female* (mean = 42.05) than *Neutral* (mean = 27.30; $t = 3.02$, $p = 0.01$). Thus, participants appeared to view our context as male-dominated or assume the subordinate was male.

¹⁶ Planned comparisons in Table 3, Panel C show no significant difference in *Potential* in *Male* (mean = 4.41) and *Neutral* (mean = 4.57; $t = -0.34$, $p = 0.74$), but *Potential* was significantly higher in *Female* (mean = 5.35) than *Neutral* (mean = 4.57; $t = 2.05$, $p = 0.05$).

TABLE 2
H1a and H1b: Supervisors' Causal Attributions of Past Performance

Panel A: Mean (Standard Deviation) of Causal Attributions of Past Performance

Points Allocated to Causal Factor ^c	<i>Strong Past Performance^a</i>			<i>Weak Past Performance^a</i>		
	<i>Gender^b</i>			<i>Gender^b</i>		
	<i>Male</i> n = 22	<i>Female</i> n = 23	<i>Neutral</i> n = 21	<i>Male</i> n = 25	<i>Female</i> n = 21	<i>Neutral</i> n = 23
<i>Ability (or lack thereof)</i>	48.41 (17.21)	36.35 (12.54)	49.52 (14.13)	25.80 (13.59)	42.05 (19.38)	27.30 (12.55)
<i>Effort</i>	34.55 (15.73)	38.13 (13.62)	33.10 (12.99)	25.00 (14.72)	28.00 (11.94)	21.87 (14.12)
<i>Luck</i>	7.73 (7.68)	15.83 (17.70)	8.81 (10.11)	17.60 (17.21)	19.24 (13.85)	18.22 (18.21)
<i>Unreasonable Target</i>	7.95 (10.76)	9.26 (8.33)	7.86 (14.80)	31.20 (28.99)	10.71 (11.49)	31.52 (27.57)
<i>Other</i>	1.36 (4.67)	0.43 (2.09)	0.71 (2.39)	0.40 (2.00)	0.00 (0.00)	1.09 (5.21)

Panel B: H1a ANOVA for Effect of Gender on Ability in Strong Past Performance^a

Source of Variation	df	Mean Square	F	p
<i>Gender^b</i>	2	1,197.24	5.52	0.01
Error	63	217.04		
Total	65			

Panel C: H1a Planned Comparisons for Ability in Strong Past Performance^a

<i>Gender^b</i>	Difference	Std. Err.	t	df	p (two-tailed)
1) <i>Female</i> versus <i>Male</i>	(12.06)	4.47	(2.70)	43	0.01
2) <i>Male</i> versus <i>Neutral</i>	(1.11)	4.82	(0.23)	41	0.82
3) <i>Female</i> versus <i>Neutral</i>	(13.17)	4.02	(3.28)	42	< 0.01

Panel D: H1b ANOVA for Effect of Gender on Ability in Weak Past Performance^a

Source of Variation	df	Mean Square	F	p
<i>Gender^b</i>	2	1,774.50	7.60	0.01
Error	66	233.51		
Total	68			

Panel E: H1b Planned Comparisons for Ability in Weak Past Performance^a

<i>Gender^b</i>	Difference	Std. Err.	t	df	p (two-tailed)
1) <i>Female</i> versus <i>Male</i>	16.25	4.88	3.33	44	0.01
2) <i>Male</i> versus <i>Neutral</i>	(1.50)	3.79	(0.40)	46	0.70
3) <i>Female</i> versus <i>Neutral</i>	14.74	4.88	3.02	42	0.01

^a We manipulated *Past Performance* by providing actual performance and targets for five quantitative performance measures, above- or below-par ratings for each measure, performance notes, and an overall above- or below-par performance rating. *Strong (Weak)* past performance was operationalized as a majority of above-par (below-par) ratings for the five quantitative performance measures and an overall above-par (below-par) rating.

^b We manipulated *Gender* by including male, female, or gender-neutral wording and silhouettes just before the subordinate's past performance information, which was either *Strong* or *Weak*.

^c We measured causal attributions of the subordinate's past performance by asking participants to allocate 100 points across five factors that could have driven past performance, with more points indicating greater influence.

TABLE 3

H2a: Supervisors' Appraisals of a Subordinate's Potential

Panel A: Mean (Standard Deviation) of *Potential*^a

	<i>Strong Past Performance</i> ^b			<i>Weak Past Performance</i> ^b		
	<i>Gender</i> ^c			<i>Gender</i> ^c		
	<i>Male</i> n = 22	<i>Female</i> n = 23	<i>Neutral</i> n = 21	<i>Male</i> n = 25	<i>Female</i> n = 21	<i>Neutral</i> n = 23
<i>Potential</i> ^a	4.41 (1.50)	5.35 (0.78)	4.57 (1.63)	3.48 (1.05)	3.71 (0.78)	3.35 (1.11)

Panel B: ANOVA for Effect of *Gender*^c on *Potential*^a in *Strong Past Performance*^b

Source of Variation	df	Mean Square	F	p
<i>Gender</i> ^c	2	5.68	3.15	0.05
Error	63	1.80		
Total	65			

Panel C: Planned Comparisons for *Potential*^a in *Strong Past Performance*^a

<i>Gender</i> ^c	Difference	Std. Err.	t	df	p (two-tailed)
1) <i>Female</i> versus <i>Male</i>	0.94	0.35	2.65	43	0.01
2) <i>Male</i> versus <i>Neutral</i>	(0.16)	0.48	(0.34)	41	0.74
3) <i>Female</i> versus <i>Neutral</i>	0.78	0.38	2.05	42	0.05

Panel D: ANOVA for Effect of *Gender*^c on *Potential*^a in *Weak Past Performance*^b

Source of Variation	df	Mean Square	F	p
<i>Gender</i> ^c	2	0.75	0.76	0.47
Error	66	1.00		
Total	68			

^a We measured participants' appraisals of the subordinate's potential with this question, anchored on 1 = Strongly disagree and 7 = Strongly agree: "To what extent do you agree that [Thomas Roan/Jennifer Roan/the Portfolio Manager] has what it takes to succeed if [he/she/he or she] is promoted to the next highest level within the bank?"

^b We manipulated *Past Performance* by providing actual performance and targets for five quantitative performance measures, above- or below-par ratings for each measure, performance notes, and an overall above- or below-par performance rating. *Strong* (*Weak*) past performance was operationalized as a majority of above-par (below-par) ratings for the five quantitative performance measures and an overall above-par (below-par) rating.

^c We manipulated *Gender* by including male, female, or gender-neutral wording and silhouettes just before the subordinate's past performance information, which was either *Strong* or *Weak*.

Table 3, Panel D, reports the ANOVA for *Weak*, which finds that *Potential* does not differ across *Gender* ($F_{2,68} = 0.76$, $p = 0.47$). That is, despite viewing a weak-performing female as having lower ability than an identically performing male, participants appraised the female subordinate's potential (mean = 3.71) as being the same as the male subordinate's (mean = 3.48).

Taken together, results support H2a. Regardless of whether past performance is strong or weak, supervisors rate female subordinates' potential higher than would be expected given underlying beliefs about ability.

H2b

H2b predicts that supervisors' beliefs about a subordinate's ability will have less influence on appraisals of potential when the subordinate is female rather than male. We tested H2b using this regression model and the *Male* (coded 0) and *Female* (coded 1) conditions:

$$Potential = \beta_0 + \beta_1(Ability) + \beta_2(Gender) + \beta_3(Ability \times Gender) + \beta_4(Overlap) + \varepsilon \quad (1)$$

We included *Overlap* as a control, since (holding past performance constant) the relationship between attributions of performance to ability and the potential to succeed in a higher-level position should account for beliefs about the overlap in skills required for both positions.¹⁷

Results for *Strong* are in Table 4, Panel A. The *Ability* \times *Gender* interaction was significant ($\beta_3 = -0.05$, $p = 0.03$), indicating that attributions to ability exerted less influence on supervisors' appraisals of potential when they appraised a female rather than a male subordinate. When Model (1) was run separately for each *Gender* condition in *Strong*, results (Table 4, Panel B) showed that *Ability* was a significant predictor of *Potential* in both *Male* and *Neutral* (for both models, $p = 0.01$), but not in *Female* ($p = 0.72$).

Results for *Weak* are in Table 4, Panel C. The (lack of) *Ability* \times *Gender* interaction is significant ($\beta_3 = 0.45$, $p = 0.01$). When Model (1) was run separately for each *Gender* condition in *Weak*, results (Table 4, Panel D) showed *Ability* was a significant predictor of *Potential* in both *Male* and *Neutral* ($p = 0.06$ and $p = 0.02$, respectively), but not in *Female* ($p = 0.34$).

Overall, results support H2b. As predicted, supervisors' beliefs about a subordinate's ability have less influence on appraisals of potential for female than male subordinates.¹⁸

Supplemental Analysis: Alternative Explanation for Pattern of Results

It is important to rule out an alternative explanation for our pattern of results that may also derive from gender stereotypes—that participants believed analytic skills are required for the current but not the higher-level position but communal skills are required for the higher-level but not the current position. Because males and females are stereotypically believed to have superior analytic and communal skills, respectively (Heilman 2012; Abele and Wojciszke 2014; Vial and Napier 2018), attributions to ability derived from performance in the current position would be lower for females than males (as in H1a and H1b), but appraisals of potential for the higher-level position would be higher for females than males (H2a). We first note that this pattern indeed holds in the *Strong* condition, but in tests of H2a for *Weak*, females' potential is directionally higher but statistically the same as, not higher than, males'. We next rely on features of our experimental design, prior research, and an analysis of participants' beliefs about the overlap in the skills required for the two positions to document why this explanation is unlikely to explain our results (Asay, Guggenmas, Kadous, Koonce, and Libby 2021).

First, we described job responsibilities for both the Portfolio Manager (current) and Team Leader (higher-level) positions that require communal skills (working with current and attracting new clients; managing subordinates) and analytic skills (managing clients' wealth and estates; managing team resources). Given that, we would expect that stereotypic beliefs that females have superior communal skills would benefit female subordinates both when supervisors are appraising the drivers of performance in the current role (i.e., if females are expected to perform well in roles that require communal skills and supervisors attribute performance in a manner consistent with their expectations, supervisors should have been more likely to attribute females' performance to ability) and their potential to succeed in the higher level role. Further, prior research finds that females seeking leadership roles in male-dominated contexts do not significantly benefit from the stereotypical assumption that they have stronger communal skills than their male counterparts (Mölders, Brosi, Bekk, Spörle, and Welp 2018; Vial and Napier 2018).

Second, we purposely used the word *Leader* in the title for the higher-level position. Both men and women stereotypically associate leadership roles, especially those in male-dominated industries, with males (Heilman 2001; Schein and Davidson 1993; Schein 2007; Johnson et al. 2008; Koenig et al. 2011; Badura et al. 2018; Offermann and Coats 2018; Fuhrmans 2020). As such, priming participants with the word *Leader* biases against finding that supervisors appraise females' potential as higher than males' in H2a because they assume a "leader" position is better suited for females' skills.

Third, the analyses of *Overlap* reported earlier indicate that the position descriptions prompted participants to believe there was at least some overlap in the skills required for the two positions, as intended, and those beliefs did not differ across *Gender*

¹⁷ All statistical inferences remain unchanged when we exclude *Overlap* from our regression analyses, or when we include an *Overlap* \times *Ability* interaction that is not significant (all $p > 0.20$). Our finding that *Overlap* is not a significant predictor of potential is consistent with prior research finding that when faced with the difficult and subjective task of predicting a subordinate's potential for success in a role s/he has not yet performed, managers tend to substitute the answer to the related but easier question of whether the subordinate is successful in his or her current role (Cricher and Rosenzweig 2014; Benson et al. 2019).

¹⁸ We also conducted separate path analyses for *Strong* and *Weak* using maximum likelihood estimation in AMOS. For each, results confirm that subordinates' gender directly influences both supervisors' appraisals of potential and attributions of subordinates' past performance to ability; and attributions to ability influence potential, but do so to a lesser extent for female than male subordinates (in both models, all four paths $p < 0.05$ in the predicted directions). However, results should be interpreted with caution; our sample size-to-parameters ratios of 16.5:1 and 17.25:1 in *Strong* and *Weak*, respectively, are smaller than a 20:1 guideline, and even higher sample sizes are suggested for models with non-continuous variables (Kline 2016, 15–16, 459).

TABLE 4

H2b: Influence of Subordinate's Ability and Gender on Appraisals of Potential

Panel A: Regression with Male and Female Conditions in Strong Past Performance^a

Variable	β	p (two-tailed)
Intercept	1.80	0.05
Ability ^b	0.05	<0.01
Gender ^c	3.36	<0.01
Ability \times Gender	(0.05)	0.03
Overlap ^d	0.08	0.57
n		45
R ²		0.34
Regression		
Sum of Squares		24.25
df		4
Mean Square		6.06
F		5.25
p		<0.01
Residual		
Sum of Squares		46.19
df		40
Mean Square		1.16
Total		
Sum of Squares		70.44
df		44

$$Potential = \beta_0 + \beta_1(Ability) + \beta_2(Gender) + \beta_3(Ability \times Gender) + \beta_4(Overlap) + \varepsilon$$

See note e (at the end of the table notes) for measure of *Potential*.

Panel B: Regressions for Each Gender Condition in Strong Past Performance^a

Variable	Gender ^c					
	Male		Female		Neutral	
	β	p (two-tailed)	β	p (two-tailed)	β	p (two-tailed)
Intercept	1.93	0.12	4.92	<0.01	2.57	0.24
Ability ^b	0.05	0.01	-0.01	0.72	0.06	0.01
Overlap ^d	0.05	0.83	0.15	0.43	-0.29	0.37
n		22		23		21
R ²		0.30		0.03		0.44
Regression						
Sum of Squares		14.02		0.45		23.27
df		2		2		2
Mean Square		7.01		0.23		11.64
F		4.00		0.36		7.01
p		0.04		0.70		<0.01
Residual						
Sum of Squares		33.30		12.76		29.87
df		19		20		18
Mean Square		1.75		0.64		1.66
Total						
Sum of Squares		47.32		13.22		53.14
df		21		22		20

$$Potential = \beta_0 + \beta_1(Ability) + \beta_2(Overlap) + \varepsilon$$

See note e (at the end of the table notes) for measure of *Potential*.

(continued on next page)

TABLE 4 (continued)

Panel C: Regression with Male and Female Conditions in Weak Past Performance^a

Variable	β	p (two-tailed)
Intercept	3.89	<0.01
Ability ^b	-0.04	0.01
Gender ^c	-1.05	0.09
Ability \times Gender	0.45	0.01
Overlap ^d	0.16	0.30
n		46
R ²		0.19
Regression		
Sum of Squares		7.52
df		4
Mean Square		1.88
F		2.44
p		<0.06
Residual		
Sum of Squares		31.64
df		41
Mean Square		0.77
Total		
Sum of Squares		39.15
df		45

$Potential = \beta_0 + \beta_1(Ability) + \beta_2(Gender) + \beta_3(Ability \times Gender) + \beta_4(Overlap) + \varepsilon$
 See note e (at the end of the table notes) for measure of *Potential*.

Panel D: Regressions for Each Gender Condition in Weak Past Performance^a

Variable	Gender ^c					
	Male		Female		Neutral	
	β	p (two-tailed)	β	p (two-tailed)	β	p (two-tailed)
Intercept	4.06	<0.01	2.66	<0.01	4.48	<0.01
Ability ^b	-0.04	0.02	0.01	0.40	-0.05	0.01
Overlap ^d	0.10	0.64	0.22	0.28	0.05	0.77
n		25		21		23
R ²		0.21		0.11		0.32
Regression						
Sum of Squares		5.60		1.40		8.56
df		2		2		2
Mean Square		2.80		0.70		4.28
F		2.98		1.16		4.59
p		0.06		0.34		0.02
Residual						
Sum of Squares		20.64		10.88		18.65
df		22		18		20
Mean Square		0.94		0.61		0.93
Total						
Sum of Squares		26.24		12.29		27.22
df		24		20		22

$Potential = \beta_0 + \beta_1(Ability) + \beta_2(Overlap) + \varepsilon$
 See note e (at the end of the table notes) for measure of *Potential*.

(continued on next page)

TABLE 4 (continued)

- ^a We manipulated *Past Performance* by providing actual performance and targets for five quantitative performance measures, above- or below-par ratings for each measure, performance notes, and an overall above- or below-par performance rating. *Strong (Weak)* past performance was operationalized as a majority of above-par (below-par) ratings for the five quantitative performance measures and an overall above-par (below-par) rating.
- ^b We measured causal attributions of the subordinate's past performance by asking participants to allocate 100 points across five factors that could have driven past performance, with more points indicating greater influence. The factors were *ability, effort, luck, targets not set at the right level*, and *other* (with an open-ended response to explain).
- ^c We manipulated *Gender* by including male or female wording and silhouettes just before the subordinate's past performance information, which was either *Strong* or *Weak*. In this analysis, *Gender* was coded 0 = male and 1 = female.
- ^d We measured participants' beliefs about the extent of overlap in the skills required to be an effective portfolio manager and team leader with this question, anchored on 1 = No overlap and 7 = Complete overlap: "To what extent do the skills required to be an effective portfolio manager overlap with the skills necessary to be an effective team leader?"
- ^e We measured participants' appraisals of the subordinate's potential with this question, anchored on 1 = Strongly disagree and 7 = Strongly agree: "To what extent do you agree that [Thomas Roan/Jennifer Roan/the Portfolio Manager] has what it takes to succeed if [he/she/he or she] is promoted to the next highest level within the bank?"

conditions. Further, as noted with tests of H2b, when an *Overlap* \times *Ability* interaction is included in the regression models, none of the interactions are significant, indicating that the weight participants placed on their beliefs about the subordinate's ability (or lack thereof) was not dependent on beliefs about the relative importance of particular skills for the current and higher-level positions. Put another way: For the alternative explanation to drive results, as participants' beliefs about the overlap in skills required for the two positions decreases on the scale (e.g., as the divergence between beliefs about the importance of analytic skills for the current position and communal skills for the higher-level position increases), the weight placed on ability (in this case, analytic ability) when making appraisals of potential should decrease since ability in the current position is a less informative signal of potential. However, based on the non-significant interactions, it does not.

Overall, these design choices, evidence from prior literature, and analyses of participant beliefs provide some evidence that our pattern of results is not due to stereotypical beliefs about females' and males' communal and analytic skills and their relevance to each position, but rather to our theory about conflicting signals sent by today's societal and organizational focus on improving the experiences and outcomes of females in male-dominated contexts.

V. CONCLUSION

In recent years, society and organizations have become increasingly focused on improving the workplace experiences and outcomes of females working in male-dominated contexts. This study examines whether, given this context, supervisors differentially interpret and use accounting information in appraising the potential of subordinates of different genders to succeed in higher-level positions. We predict and find that signals sent by today's social and organizational context influence the use of accounting information in appraisal processes and outcomes in countervailing ways. On one hand, we find gender stereotypes persist when supervisors use past performance information to form beliefs about the drivers of that performance—females are believed to have lower ability than equally performing males. On the other hand, females appear to be valued from a diversity standpoint, in that appraisals of their potential are higher than those of males—and higher than predicted by prior accounting research, given supervisors' underlying beliefs about females' abilities. Specifically, we find that despite viewing females as less able than equally performing males, supervisors judge a female subordinate's potential to be the same as (in the case of weak performance) or higher than (in the case of strong performance) a male's potential. Further, supervisors place less weight on their underlying beliefs about subordinates' ability when appraising female subordinates' potential.

While our findings regarding appraisals of potential appear to advantage female subordinates, they suggest those ratings may not accurately capture supervisors' true underlying beliefs. Thus, today's societal and organizational context does benefit females in that supervisors appraise a strongly performing female's potential as being higher than an equally performing male's, but this benefit does not arise because gender stereotypes are less prevalent. Rather, it appears that supervisors value subordinates based not only on ability, but also on perceived contributions to diversity norms and goals. The perception that females receive high potential ratings because of their diversity value can generate hostility and animus toward females, and cause females to question their own qualifications and lose confidence and motivation (Caleo and Heilman 2019; Leslie et al. 2014).

The persistence of gender stereotypes may help explain why diversity initiatives have done little to move the needle on outcomes for females identified as high-potential (McKinsey 2020). For example, evidence suggests that high-potential females are paired with mentors who have less organizational clout and who are less likely to advocate on their mentees' behalf, and may be given fewer responsibilities that help hone competencies needed for higher-level positions (Ibarra et al. 2010; Fernández-Aráoz et al. 2017). Further, females are more likely to be promoted to leadership roles in organizations that are

failing or in crisis, reducing their chances of success (Ryan et al. 2016). When females are promoted to leadership roles, they are held to higher standards and are more likely to be blamed for failures (Heilman and Caleo 2018; Brescoll et al. 2010; Manzi and Heilman 2021). While our study examines a focus on the underrepresentation of females in male-dominated contexts, we expect our findings would similarly apply to appraisals of other underrepresented group members.

Our study has limitations that provide opportunities for additional research. First, we do not directly measure whether our observed differences are driven by a societal and organizational focus on improving females' workplace experiences and outcomes, nor do we manipulate the presence of diversity initiatives. Notably, we could not find academic or practitioner studies on supervisors' beliefs about diversity initiatives or pressures they experience when appraising females, suggesting that either others have not collected such data or have run into obstacles when doing so like those we anticipated. A worthy goal of future research is systematically gathering unbiased data on these beliefs. Second, our experimental task necessarily excludes an array of information that supervisors might want to incorporate into appraisals of potential, particularly measures of subordinates' abilities with respect to higher-level positions. Future research could ascertain what information supervisors would like to have or choose to use when appraising potential and how they use it, or how noise in measures of subordinates' abilities they have not needed to use impacts appraisals of potential. Third, the order in which we asked task-related questions could have impacted responses. Indeed, asking supervisors to consider their performance attributions or the overlap in skills required to be successful in the current and higher-level positions before forming appraisal ratings could prompt more systematic processing, serving as a debiasing mechanism for overly favorable or unfavorable appraisals for particular genders. More broadly, this study speaks to a need for researchers and organizations to weigh the implications of our results for subordinates' motivation, job satisfaction, and productivity. Organizations would be well served by research examining whether the design of their high-potential and diversity initiatives in fact help them achieve their goals. As examples, many organizations use diversity training as a personnel control, but its value has been questioned; there are also calls to reduce ambiguity in appraisal processes, and to use cultural controls to increase social accountability for advancing diversity (Dobbin and Kalev 2016; Mackenzie, Wehner and Correll 2019; Chamorro-Premuzic 2020; McKinsey 2020; Ro 2021).

Our research contributes to a growing body of accounting literature focusing on appraisals of subordinates' potential to succeed in different, higher-level positions. In addition, few accounting studies on appraisals consider the impact of subordinate gender on appraisal processes and outcomes, but understanding whether and how the salience of gender in today's social environment affects the use of accounting information sheds light on how we can design more effective performance measurement and control systems. Importantly, our results suggest that research on the effectiveness of performance measurement and control systems could benefit from examining the differential impacts of the gender of stakeholders in these systems.

REFERENCES

- Abele, A., and B. Wojciszke. 2014. Chapter Four—Communal and Agentic Content in Social Cognition: A Dual Perspective Model. *Advances In Experimental Social Psychology* 50 (2014): 195–255. <https://doi.org/10.1016/B978-0-12-800284-1.00004-7>
- Adler, S., M. Campion, A. Colquitt, A. Grubb, K. Murphy, R. Ollander-Krane, and E. Pulakos. 2016. Getting rid of performance ratings: Genius or folly? A debate. *Industrial and Organizational Psychology: Perspectives on Science and Practice* 9 (2): 219–252. <https://doi.org/10.1017/iop.2015.106>
- Alden, W. 2014. Wall Street's young bankers are still mostly white and male, report says. *The New York Times*, September 30. Available at: <https://dealbook.nytimes.com/2014/09/30/wall-streets-young-bankers-are-still-mostly-white-and-male/>
- Anderson, S., and A. Lillis. 2011. Corporate frugality: Theory, measurement and practice. *Contemporary Accounting Research* 28 (4): 1349–1387. <https://doi.org/10.1111/j.1911-3846.2011.01107.x>
- Aon Hewitt. 2013. *Building the Right High Potential Pool: How Organizations Define, Assess, and Calibrate Their Critical Talent*. London, U.K.: Aon Hewitt.
- Asay, J., R. Guggenmas, K. Kadous., L. Koonce, and R. Libby. 2021. *Theory testing and process evidence in accounting experiments*. Working paper, University of Iowa, Cornell University, Emory University, and University of Texas at Austin. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3485844
- Badura, K., E. Grijalva, D. Newman, T. Yan, and G. Jeon. 2018. Gender and leadership emergence: A meta-analysis and explanatory model. *Personnel Psychology* 71 (3): 335–367. <https://doi.org/10.1111/peps.12266>
- Banker, R., and S. Datar. 1989. Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research* 27 (1): 21–39. <https://doi.org/10.2307/2491205>
- Benson, A., D. Li, and K. Shue. 2019. Promotions and the Peter Principle. *The Quarterly Journal of Economics* 134 (4): 2085–2134. <https://doi.org/10.1093/qje/qjz022>
- Biernat, M., and K. Fiegen. 2001. Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *The Journal of Social Issues* 57 (4): 707–724. <https://doi.org/10.1111/0022-4537.00237>

- Bigelow, L., L. Lundmark, J. McLean Parks, and R. Wuebker. 2014. Skirting the issues: Experimental evidence of gender bias in IPO prospectus evaluations. *Journal of Management* 40 (6): 1732–1759. <https://doi.org/10.1177/0149206312441624>
- Birnberg, J., I. Frieze, and M. Shields. 1977. The role of attribution theory in control systems. *Accounting, Organizations and Society* 2 (3): 189–200. [https://doi.org/10.1016/0361-3682\(77\)90011-3](https://doi.org/10.1016/0361-3682(77)90011-3)
- Blau, F. D., and L. M. Kahn. 2017. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* 55 (3): 789–865. <https://doi.org/10.1257/jel.20160995>
- Block, K., A. Croft, L. De Souza, and T. Schmader. 2019. Do people care if men don't care about caring? The asymmetry in support for changing gender roles. *Journal of Experimental Social Psychology* 83: 112–131. <https://doi.org/10.1016/j.jesp.2019.03.013>
- Bloomfield, R., K. Rennekamp, B. Steenhoven, and S. Stewart. 2021. Penalties for unexpected behavior: Double standards for women in finance. *The Accounting Review* 96 (2): 107–125. <https://doi.org/10.2308/tar-2018-0715>
- Bol, J. 2008. Subjectivity in compensation contracting. *Journal of Accounting Literature* 27: 1–32.
- Bol, J., and J. Leiby. 2018. Subjectivity in professionals' incentive systems: Differences between promotion- and performance-based assessments. *Contemporary Accounting Research* 35 (1): 31–57. <https://doi.org/10.1111/1911-3846.12393>
- Bol, J., and S. Smith. 2011. Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review* 86 (4): 1213–1230. <https://doi.org/10.2308/accr-10038>
- Bowen, D., and C. Ostroff. 2004. Understanding HRM-firm performance linkages: The role of the “strength” of the HRM system. *Academy of Management Review* 29 (2): 203–221. <https://doi.org/10.5465/amr.2004.12736076>
- Braun, S., S. Stegmann, A. Hernandez Bark, N. Junker, and R. van Dick. 2017. Think manager—think male, think follower—think female: Gender bias in implicit followership theories. *Journal of Applied Social Psychology* 47 (7): 377–388. <https://doi.org/10.1111/jasp.12445>
- Brescoll, V., E. Dawson, and E. Uhlmann. 2010. Hard won and easily lost: The fragile status of leaders in gender-stereotype-incongruent occupations. *Psychological Science* 21 (11): 1640–1642. <https://doi.org/10.1177/0956797610384744>
- Brewer, M. 1988. A dual process model of impression formation. In *Advances in Social Cognition*, edited by T. Srull and R. Wyer, 1–36. Hillsdale, NJ: Lawrence Erlbaum.
- Brosnan, A. 2018. *Women in business: Beyond policy to progress*. Grant Thornton International Ltd. Available at: <https://www.granthornton.global/globalassets/1.-member-firms/global/insights/women-in-business/grant-thornton-women-in-business-2018-report.pdf>
- Caleo, S., and M. Heilman. 2019. What could go wrong? Some unintended consequences of gender bias interventions. *Archives of Scientific Psychology* 7 (1): 71–80. <https://doi.org/10.1037/arc0000063>
- Cappelli, P., and J. Keller. 2014. Talent management: Conceptual approaches and practical challenges. *Annual Review of Organizational Psychology and Organizational Behavior* 1 (1): 305–331. <https://doi.org/10.1146/annurev-orgpsych-031413-091314>
- Catalyst. 2018. *Quick take: Women in management*. Available at: <https://www.catalyst.org/research/women-in-management/>
- Cejka, M., and A. Eagly. 1999. Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin* 25 (4): 413–423. <https://doi.org/10.1177/0146167299025004002>
- Chamorro-Premuzic, T. 2020. *Science explains why unconscious bias training won't reduce workplace racism. Here's what will*. Fast Company. Available at: <https://www.fastcompany.com/90515678/science-explains-why-unconscious-bias-training-wont-reduce-workplace-racism-heres-what-will>
- Chamorro-Premuzic, T., S. Adler and R. Kaiser. 2017. What science says about identifying high-potential employees. *Harvard Business Review*. Available at: <https://hbr.org/2017/10/what-science-says-about-identifying-high-potential-employees>
- Chan, E. 2018. Promotion, relative performance information, and the Peter Principle. *The Accounting Review* 93 (3): 83–103. <https://doi.org/10.2308/accr-51890>
- Church, A., and C. Rotolo. 2013. How are top companies assessing their high-potentials and senior executives? A talent management benchmark study. *Consulting Psychology Journal* 65 (3): 199–223. <https://doi.org/10.1037/a0034381>
- Church, A. H., C. T. Rotolo, N. M. Ginther, and R. Levine. 2015. How are top companies designing and managing their high-potential programs? A follow-up talent management benchmark study. *Consulting Psychology Journal* 67 (1): 17–47. <https://doi.org/10.1037/cpb0000030>
- Critcher, C., and E. Rosenzweig. 2014. The performance heuristic: A misguided reliance on past success when predicted prospects for improvement. *Journal of Experimental Psychology: General* 143 (2): 480–485. <https://doi.org/10.1037/a0034129>
- Deaux, K., and T. Emswiller. 1974. Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology* 29 (1): 80–85. <https://doi.org/10.1037/h0035733>
- Demeré, B., K. Sedatole, and A. Woods. 2019. The role of calibration committees in subjective performance evaluation systems. *Management Science* 65 (4): 1562–1585. <https://doi.org/10.1287/mnsc.2017.3025>
- DeNisi, A., T. Cafferty, and B. Meglino. 1984. A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance* 33 (3): 360–396. [https://doi.org/10.1016/0030-5073\(84\)90029-1](https://doi.org/10.1016/0030-5073(84)90029-1)
- Dichev, I., J. Graham, C. Harvey, and S. Rajgopal. 2013. Earnings quality: Evidence from the field. *Journal of Accounting and Economics* 56 (2–3 Supplement 1): 1–33. <https://doi.org/10.1016/j.jacceco.2013.05.004>
- Dobbin, F., and A. Kalev. 2016. Why diversity programs fail. *Harvard Business Review* 94 (7/8): 52–60.

- Dover, T. L., C. Kaiser, and B. Major. 2020. Mixed signals: The unintended effects of diversity initiatives. *Social Issues and Policy Review* 14 (1): 152–181. <https://doi.org/10.1111/sipr.12059>
- Fairburn, J., and J. Malcomson. 2001. Performance, promotion, and the Peter Principle. *The Review of Economic Studies* 68 (1): 45–66. <https://doi.org/10.1111/1467-937X.00159>
- Feldman, J. 1981. Beyond attribution theory: Cognitive processes in performance appraisal. *The Journal of Applied Psychology* 66 (2): 127–148. <https://doi.org/10.1037/0021-9010.66.2.127>
- Feldman-Summers, S., and S. Kiesler. 1974. Those who are number two try harder: The effect of sex on attributions of causality. *Journal of Personality and Social Psychology* 30 (6): 846–855. <https://doi.org/10.1037/h0037604>
- Feltham, G., and J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69 (3): 429–453.
- Fernández-Aráoz, C., A. Roscoe, and K. Aramaki. 2017. Turning potential into success: The missing link in leadership development. *Harvard Business Review* 95 (6): 86–93.
- Finkelstein, L., D. Costanza, and G. Goodwin. 2018. Do your high potentials have potential? The impact of individual differences and designation on leader success. *Personnel Psychology* 71 (1): 3–22. <https://doi.org/10.1111/peps.12225>
- Foschi, M. 2000. Double standards for competence: Theory and research. *Annual Review of Sociology* 26 (1): 21–42. <https://doi.org/10.1146/annurev.soc.26.1.21>
- Fuhrmans, V. 2020. Where are all the women CEOs? *Wall Street Journal*, February 6. Available at: <https://www.wsj.com/articles/why-so-few-ceos-are-women-you-can-have-a-seat-at-the-table-and-not-be-a-player-11581003276>
- Grabner, I., and F. Moers. 2013. Managers' choices of performance measures in promotion decisions: An analysis of alternative job assignments. *Journal of Accounting Research* 51 (5): 1187–1220. <https://doi.org/10.1111/1475-679X.12027>
- Graham, J., C. Harvey, and M. Puri. 2013. Managerial attitudes and corporate actions. *Journal of Financial Economics* 109 (1): 103–121. <https://doi.org/10.1016/j.jfineco.2013.01.010>
- Green, S., and T. Mitchell. 1979. Attributional processes of leaders in leader-member interactions. *Organizational Behavior and Human Performance* 23 (3): 429–458. [https://doi.org/10.1016/0030-5073\(79\)90008-4](https://doi.org/10.1016/0030-5073(79)90008-4)
- Greenhaus, J., and S. Parasuraman. 1993. Job performance attributions and career advancement prospects: An examination of gender and race effects. *Organizational Behavior and Human Decision Processes* 55 (2): 273–297. <https://doi.org/10.1006/obhd.1993.1034>
- Harrison, P., S. West, and J. Reneau. 1988. Initial attributions and information-seeking by superiors and subordinates in production variance investigations. *The Accounting Review* 63 (2): 307–320.
- Harvey, P., K. Madison, M. Martinko, T. R. Crook, and T. A. Crook. 2014. Attribution theory in the organizational sciences: The road traveled and the path ahead. *The Academy of Management Perspectives* 28 (2): 128–146. <https://doi.org/10.5465/amp.2012.0175>
- Hegewisch, A., and E. Mefferd. 2021. *The gender wage gap by occupation, race, and ethnicity 2020*. IWPR Policy Brief #C497. Institute for Women's Policy Research, March. Available at: <https://iwpr.org/wp-content/uploads/2021/03/2021-Occupational-Wage-Gap-Brief-v2.pdf>
- Heider, F. 1958. *The Psychology of Interpersonal Relations*. Hoboken, NJ: John Wiley and Sons, Inc.
- Heilman, M. 1983. Sex bias in work settings: The lack of fit model. In *Research in Organizational Behavior*. Volume 5, edited by B. Staw and L. Cummings, 269–298. Greenwich, CT: JAI Press.
- Heilman, M. 2001. Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *The Journal of Social Issues* 57 (4): 657–674. <https://doi.org/10.1111/0022-4537.00234>
- Heilman, M. 2012. Gender stereotypes and workplace bias. *Research in Organizational Behavior* 32: 113–135. <https://doi.org/10.1016/j.riob.2012.11.003>
- Heilman, M., and S. Caleo. 2018. Combatting gender discrimination: A lack of fit framework. *Group Processes & Intergroup Relations* 21 (5): 725–744. <https://doi.org/10.1177/1368430218761587>
- Heilman, M., and R. Guzzo. 1978. The perceived cause of work success as a mediator of sex discrimination in organizations. *Organizational Behavior and Human Performance* 21 (3): 346–357. [https://doi.org/10.1016/0030-5073\(78\)90058-2](https://doi.org/10.1016/0030-5073(78)90058-2)
- Heilman, M., C. Block, R. Martell, and M. Simon. 1989. Has anything changed? Current characterizations of men, women, and managers. *The Journal of Applied Psychology* 74 (6): 935–942. <https://doi.org/10.1037/0021-9010.74.6.935>
- Hentschel, T., M. Heilman, and C. Peus. 2019. The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves. *Frontiers in Psychology* 10: 11. <https://doi.org/10.3389/fpsyg.2019.00011>
- Holman, J., L. Keller, and L. Colby. 2018. Banks aren't showing their math on identical 1% gender pay gaps. *Bloomberg*, February 6. Available at: <https://www.bloomberg.com/news/articles/2018-02-06/four-wall-street-banks-all-reveal-a-1-pay-gap-coincidence>
- Ibarra, H., N. Carter, and C. Silva. 2010. Why men still get more promotions than women. *Harvard Business Review* 88 (9): 80–85.
- Ilgén, D., J. Barnes-Farrell, and D. McKellin. 1993. Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes* 54 (3): 321–368. <https://doi.org/10.1006/obhd.1993.1015>
- Ilgén, D., and J. Feldman. 1983. Performance appraisal: A process focus. In *Research in Organizational Behavior*. Volume 5, edited by B. Staw and L. Cummings, 141–197. Greenwich, CT: JAI Press.
- Jaekel, A., and E. St-Onge. 2016. Why women aren't making it to the top of financial services firms. *Harvard Business Review*. Available at: <https://hbr.org/2016/10/why-women-arent-making-it-to-the-top-of-financial-services-firms>

- Johns, G. 2006. The essential impact of context on organizational behavior. *Academy of Management Review* 31 (2): 386–408. <https://doi.org/10.5465/amr.2006.20208687>
- Johnson, C., and K. Hawbaker. 2019. #MeToo: A timeline of events. *Chicago Tribune*, July 10. Available at: <https://www.chicagotribune.com/lifestyles/ct-me-too-timeline-20171208-htmlstory.html>
- Johnson, S., S. Murphy, S. Zewdie, and R. Reichard. 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 (1): 39–60. <https://doi.org/10.1016/j.obhdp.2007.12.002>
- Kaplan, S., and P. Reckers. 1985. An examination of auditor performance evaluation. *The Accounting Review* 60 (3): 477–487.
- Kelley, H. 1973. The processes of causal attribution. *The American Psychologist* 28 (2): 107–128. <https://doi.org/10.1037/h0034225>
- Kline, R. 2016. *Principles and Practice of Structural Equation Modeling*. 4th edition. New York, NY: The Guilford Press.
- Koenig, A., A. Eagly, A. Mitchell, and T. Ristikari. 2011. Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin* 137 (4): 616–642. <https://doi.org/10.1037/a0023557>
- Krentz, M. 2019. Survey: What diversity and inclusion policies do employees actually want? *Harvard Business Review*. Available at: <https://hbr.org/2019/02/survey-what-diversity-and-inclusion-policies-do-employees-actually-want>
- Lambert, R. 2001. Contracting theory and accounting. *Journal of Accounting and Economics* 32 (1–3): 3–87. [https://doi.org/10.1016/S0165-4101\(01\)00037-4](https://doi.org/10.1016/S0165-4101(01)00037-4)
- Lee, L., and G. Waddell. 2021. Diversity and the timing of preference in hiring decisions. *Journal of Economic Behavior & Organization* 184: 432–459. <https://doi.org/10.1016/j.jebo.2020.11.014>
- Leslie, L. 2019. Diversity initiative effectiveness: A typological theory of unintended consequences. *Academy of Management Review* 44 (3): 538–563. <https://doi.org/10.5465/amr.2017.0087>
- Leslie, L., D. Mayer, and D. Kravitz. 2014. The stigma of affirmative action: A stereotyping-based theory and meta-analytic test of the consequences for performance. *Academy of Management Journal* 57 (4): 964–989. <https://doi.org/10.5465/amj.2011.0940> <https://psycnet.apa.org/doi/10.5465/amj.2011.0940>
- Leslie, L., C. Manchester, and P. Dahm. 2017. Why and when does the gender gap reverse? Diversity goals and the pay premium for high potential women. *Academy of Management Journal* 60 (2): 402–432. <https://doi.org/10.5465/amj.2015.0195>
- Levanon, A., and D. Grusky. 2016. The persistence of extreme gender segregation in the twenty-first century. *American Journal of Sociology* 122 (2): 573–619. <https://doi.org/10.1086/688628>
- Lowe, D., P. Reckers, and D. Sanders. 2001. The influence of gender, ethnicity, and individual differences on perceptions of career progression in public accounting. *International Journal of Auditing* 5 (1): 53–71. <https://doi.org/10.1111/1099-1123.00325>
- Luft, J., M. Shields, and T. Thomas. 2016. Additional information in accounting reports: Effects on management decisions and subjective performance evaluations under causal ambiguity. *Contemporary Accounting Research* 33 (2): 526–550. <https://doi.org/10.1111/1911-3846.12156>
- Lyness, K., and M. Heilman. 2006. When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *The Journal of Applied Psychology* 91 (4): 777–785. <https://doi.org/10.1037/0021-9010.91.4.777>
- Maas, V., and R. Torres-González. 2011. Subjective performance evaluation and gender discrimination. *Journal of Business Ethics* 101 (4): 667–681. <https://doi.org/10.1007/s10551-011-0763-7>
- Mackenzie, L., J. Wehner, and S. Correll. 2019. Why most performance evaluations are biased, and how to fix them. *Harvard Business Review*, January 11. Available at: <https://hbr.org/2019/01/why-most-performance-evaluations-are-biased-and-how-to-fix-them>
- MacRae, I., and A. Furnham. 2014. *High Potential: How to Spot, Manage and Develop Talented People at Work*. London, U.K.: Bloomsbury Publishing.
- Manzi, F., and M. Heilman. 2021. Breaking the glass ceiling: For one and all? *Journal of Personality and Social Psychology* 120 (2): 257–277. <https://doi.org/10.1037/pspa0000260>
- Mattone, J., and L. Xavier. 2012. *Talent Leadership: A Proven Method for Identifying and Developing High-Potential Employees*. New York, NY: AMACOM, A Division of the American Management Association.
- McKinsey. 2017. *Women in the Workplace*. New York, NY: McKinsey and Company.
- McKinsey. 2020. *Women in the Workplace*. New York, NY: McKinsey and Company.
- Mitchell, T., and R. Wood. 1980. Supervisor's responses to subordinate poor performance: A test of an attributional model. *Organizational Behavior and Human Performance* 25 (1): 123–138. [https://doi.org/10.1016/0030-5073\(80\)90029-X](https://doi.org/10.1016/0030-5073(80)90029-X)
- Moers, F. 2005. Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society* 30 (1): 67–80. <https://doi.org/10.1016/j.aos.2003.11.001>
- Mölders, S., P. Brosi, M. Bekk, M. Spörle, and I. Welp. 2018. Support for quotas for women in leadership: The influence of gender stereotypes. *Human Resource Management* 57 (4): 869–882. <https://doi.org/10.1002/hrm.21882>
- Monnery, L., and L. Blais. 2017. *To get diversity right, get potential right*. Egon Zehnder. Available at: <https://www.egonzehnder.com/insight/to-get-diversity-right-get-potential-right>
- Nederhof, A. J. 1985. Methods of coping with social desirability bias: A review. *European Journal of Social Psychology* 15 (3): 263–280. <https://doi.org/10.1002/ejsp.2420150303>
- Nieva, V., and B. A. Gutek. 1980. Sex effects on evaluation. *Academy of Management Review* 5 (2): 267–276. <https://doi.org/10.5465/amr.1980.4288749>

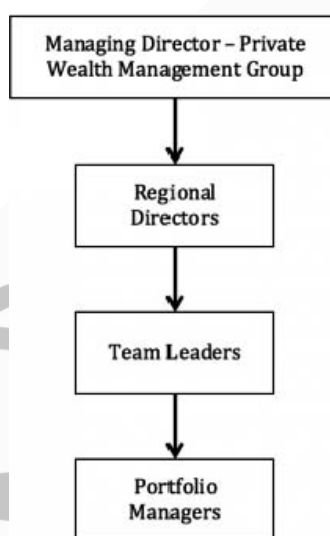
- Nisbett, R., H. Zukier, and R. Lemley. 1981. The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology* 13 (2): 248–277. [https://doi.org/10.1016/0010-0285\(81\)90010-4](https://doi.org/10.1016/0010-0285(81)90010-4)
- Offermann, L., and M. Coats. 2018. Implicit theories of leadership: Stability and change over two decades. *The Leadership Quarterly* 29 (4): 513–522. <https://doi.org/10.1016/j.leaqua.2017.12.003>
- Ostroff, C., and D. Bowen. 2000. Moving HR to a higher level: HR practices and organizational effectiveness. In *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*, edited by K. Klein and S. Kozlowski, 211–266. San Francisco, CA: Jossey-Bass.
- Paulhus, D. L., and D. B. Reid. 1991. Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology* 60 (2): 307–317. <https://doi.org/10.1037/0022-3514.60.2.307>
- Pazy, A. 1987. Sex differences in responsiveness to organizational career management. *Human Resource Management* 26 (2): 243–256. <https://doi.org/10.1002/hrm.3930260208>
- Pence, E., W. Pendleton, G. Dobbins, and J. Sgro. 1982. Effects of causal explanations and sex variables on recommendations for corrective actions following employee failure. *Organizational Behavior and Human Performance* 29 (2): 227–240. [https://doi.org/10.1016/0030-5073\(82\)90257-4](https://doi.org/10.1016/0030-5073(82)90257-4)
- Pew Research Center. 2019. *A changing world: Global views on diversity, gender equality, family life and the importance of religion*. Available at: <https://www.pewresearch.org/global/2019/04/22/a-changing-world-global-views-on-diversity-gender-equality-family-life-and-the-importance-of-religion/>
- Prendergast, C., and R. Topel. 1993. Discretion and bias in performance evaluation. *European Economic Review* 37 (2–3): 355–365. [https://doi.org/10.1016/0014-2921\(93\)90024-5](https://doi.org/10.1016/0014-2921(93)90024-5)
- PwC. 2019. Global diversity and inclusion survey. PwC Global. Available at: <https://www.pwc.com/gx/en/services/people-organisation/global-diversity-and-inclusion-survey.html#data>
- Ro, C. 2021. *Helping workers understand their implicit biases is helpful, in theory – but these programmes are controversial. Who's right?* BBC. Available at: <https://www.bbc.com/worklife/article/20210326-the-complicated-battle-over-unconscious-bias-training>
- Ryan, M., S. Haslam, T. Morgenroth, F. Rink, J. Stoker, and K. Peters. 2016. Getting on top of the glass cliff: Reviewing a decade of evidence, explanations, and impact. *The Leadership Quarterly* 27 (3): 446–455. <https://doi.org/10.1016/j.leaqua.2015.10.008>
- Sales Management Association. 2008. *Sales and sales management competencies: Designing, implementing, and maintaining competency-based management programs for the sales organization*. Available at: <https://salesmanagement.org/resource/sales-and-sales-management-competencies/>
- Scarborough, W. 2018. What the data says about women in management between 1980 and 2010. *Harvard Business Review*. Available at: <https://hbr.org/2018/02/what-the-data-says-about-women-in-management-between-1980-and-2010>
- Schein, V. 2007. Women in management: Reflections and projections. *Women in Management Review* 22 (1): 6–18. <https://doi.org/10.1108/09649420710726193>
- Schein, V., and M. Davidson. 1993. Think manager, think male. *Management Development Review* 6 (3): 24–28. <https://doi.org/10.1108/EUM00000000000738>
- Shah, A., and D. Oppenheimer. 2008. Heuristics made easy: An effort-reduction framework. *Psychological Bulletin* 134 (2): 207–222. <https://doi.org/10.1037/0033-2909.134.2.207>
- Shields, M., J. Birnberg, and I. Frieze. 1981. Attributions, cognitive processes and control systems. *Accounting, Organizations and Society* 6 (1): 69–93. [https://doi.org/10.1016/0361-3682\(81\)90023-4](https://doi.org/10.1016/0361-3682(81)90023-4)
- Shook, E., and J. Sweet. 2018. *When She Rises, We All Rise—Getting to Equal 2018: Creating a Culture Where Everyone Thrives*. Dublin, Ireland: Accenture.
- Silzer, R., and A. Church. 2009. The pearls and perils of identifying potential. *Industrial and Organizational Psychology: Perspectives on Science and Practice* 2 (4): 377–412. <https://doi.org/10.1111/j.1754-9434.2009.01163.x>
- Society for Human Resource Management. 2009. *Global Diversity and Inclusion: Perceptions, Practices and Attitudes*. Alexandria, VA: Society for Human Resource Management.
- Spence, J., and L. Keeping. 2013. The road to performance ratings is paved with intentions. *Organizational Psychology Review* 3 (4): 360–383. <https://doi.org/10.1177/2041386613485969>
- Tetlock, P., J. Lerner, and R. Boettger. 1996. The dilution effect: Judgmental bias, conversational convention, or a bit of both? *European Journal of Social Psychology* 26 (6): 915–934. [https://doi.org/10.1002/\(SICI\)1099-0992\(199611\)26:6<915::AID-EJSP797>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1099-0992(199611)26:6<915::AID-EJSP797>3.0.CO;2-W)
- Tinkler, J. 2013. How do sexual harassment policies shape gender beliefs? An exploration of the moderating effects of norm adherence and gender. *Social Science Research* 42 (5): 1269–1283. <https://doi.org/10.1016/j.ssresearch.2013.05.002>
- Tourangeau, R., and T. Yan. 2007. Sensitive questions in surveys. *Psychological Bulletin* 133 (5): 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>
- Valle, V., and I. Frieze. 1976. Stability of causal attributions as a mediator in changing expectations for success. *Journal of Personality and Social Psychology* 33 (5): 579–587. <https://doi.org/10.1037/0022-3514.33.5.579>
- Vial, A., and J. Napier. 2018. Unnecessary frills: Communitas as a nice (but expendable) trait in leaders. *Frontiers in Psychology* 9: 1866. <https://doi.org/10.3389/fpsyg.2018.01866>

- Weiner, B., and A. Kukla. 1970. An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology* 15 (1): 1–20. <https://doi.org/10.1037/h0029211>
- Weiner, B., I. Frieze, A. Kukla, L. Reed, S. Rest, and R. Rosenbaum. 1971. Perceiving the causes of success and failure. In *Attribution: Perceiving the Causes of Behavior*, edited by E. Jones, D. Kanouse, H. Kelley, R. Nisbett, S. Valins, and B. Weiner, 1–26. Morristown, NJ: General Learning Press.
- World Economic Forum. 2016. *The Industry Gender Gap: Women and Work in the Fourth Industrial Revolution*. Geneva, Switzerland: World Economic Forum.
- World Economic Forum. 2017. *The Global Gender Gap Report*. Geneva, Switzerland: World Economic Forum.
- Zarate, M., and E. Smith. 1990. Person categorization and stereotyping. *Social Cognition* 8 (2): 161–185. <https://doi.org/10.1521/soco.1990.8.2.161>
- Zillman, C. 2019. Wall Street has never had a woman CEO. Why not? *Fortune*, September 19. Available at: <https://fortune.com/longform/banking-finance-women-ceos-wall-street/>

APPENDIX A

Experiment Instrument: Overview

Grogan & Co. is a mid-size private bank. Among its many divisions is the Private Wealth Management group. The basic organizational hierarchy of this division is as follows:



Portfolio Managers work directly with clients, helping them manage their investments, income and estate planning. In addition to serving existing clients, *Portfolio Managers* are responsible for attracting new clients.

Team Leaders manage teams of 7–10 *Portfolio Managers*. They are responsible for managing resources within their team and making sure that the *Portfolio Managers* who report to them understand and act in ways consistent with the bank's strategy.

Team Leaders report to *Regional Directors* who, in addition to overseeing all *Team Leaders* within their geographic region, are responsible for developing the region's strategy and managing its resources and growth.

Regional Directors report to the *Managing Director*, who is responsible for overseeing Grogan & Co.'s entire wealth management group.

For purposes of today's session, assume that you are a *Team Leader*.

One of your responsibilities as a *Team Leader* is to evaluate the performance of the *Portfolio Managers* who report directly to you. Recently, you completed annual performance reviews for the year ended December 31, 2016.

You are glad 2016 is behind you. It was a somewhat difficult year for the entire bank because of an unexpected decline in economic conditions in the second, third, and fourth quarters.

APPENDIX B

Experiment Instrument: Gender and Past Performance Manipulations

One of your responsibilities as a Team Leader is to evaluate the performance of the Portfolio Managers who report directly to you. Recently, you completed annual performance reviews for the year ended December 31, 2016.

You are glad 2016 is behind you. It was a somewhat difficult year for the entire bank because of an unexpected decline in economic conditions in the second, third, and fourth quarters.

The Gender manipulation appears in the next line of text and accompanying silhouette as such:

Gender = Male
The performance information for one of your Portfolio Managers, Thomas Roan, is provided below.



Gender = Female
The performance information for one of your Portfolio Managers, Jennifer Roan, is provided below.



Gender = Neutral
The performance information for one of your Portfolio Managers is provided below.



The Past Performance manipulation appears in the table of performance information as such:

Past Performance = Strong

						INDIVIDUAL PERFORMANCE RATING:		
						BELOW PAR	PAR	ABOVE PAR
				INDIVIDUAL PERFORMANCE	PERFORMANCE TARGET			
Number of New Clients				6	4			X
Number of Lost Clients				1	0	X		
Increase (Decrease) in Assets U/M from Prior Year				\$ 115,997,000	\$ 116,000,000		X	
Fees Generated in Current Year				\$ 2,190,000	\$ 1,750,000			X
Client Satisfaction Score				93%	85%			X
NOTES: <i>Takes time to adapt to new systems and processes; can be easily flustered.</i> <i>Very direct and straightforward.</i> <i>Seeks advice from others when it is appropriate to do so.</i>								
OVERALL PERFORMANCE RATING								X

						INDIVIDUAL PERFORMANCE RATING:		
						BELOW PAR	PAR	ABOVE PAR
						INDIVIDUAL PERFORMANCE	PERFORMANCE TARGET	
Number of New Clients						4	6	x
Number of Lost Clients						0	1	
Increase (Decrease) in Assets U/M from Prior Year						\$ 116,000,000	\$ 115,997,000	x
Fees Generated in Current Year						\$ 1,750,000	\$ 2,190,000	x
Client Satisfaction Score						85%	93%	x
NOTES: <i>Takes time to adapt to new systems and processes; can be easily flustered.</i> <i>Very direct and straightforward.</i> <i>Seeks advice from others when it is appropriate to do so.</i>								
OVERALL PERFORMANCE RATING								x

Note that individuals' ratings (below par, par, or above par) are based on how well they perform relative to their performance targets.