



UNSW Business School

Accounting Research Seminar Series
Term 3, 2019

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

Anne Farrell

Miami University Farmer School of Business

Date: Friday Oct 18, 2019

Time: 3.00pm – 4.00pm

Venue: BUS 119

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

Anne M. Farrell
Miami University Farmer School of Business
800 E. High St.
Oxford, OH 45056, USA
Anne.Farrell@MiamiOH.edu

Michele L. Frank
Miami University Farmer School of Business
800 E. High St.
Oxford, OH 45056, USA
frankm@MiamiOH.edu

September 2019

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

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

Abstract

Initiatives to systematically appraise the extent to which employees have the potential to succeed in higher-level positions in their organizations are increasingly common. At the same time, the social context within which such appraisals take place—specifically, a norm of fostering diversity and inclusion—can induce supervisors to feel pressure to elevate high-potential employees from underrepresented groups or to avoid appearing biased against those groups. Using a setting stereotypically associated with males, we experimentally examine how supervisors use information about a subordinate's past performance when judging the subordinate's potential to succeed in a different, higher-level position, and whether the use of this information differs depending on the subordinate's gender. Relying on prior research on performance attributions and organizational socialization, we predict and find that gender has countervailing influences on appraisal processes. On one hand, supervisors are less likely to attribute strong past performance to ability when a subordinate is female rather than male, disadvantaging females. On the other hand, supervisors' appraisals of female subordinates' potential are less systematic than what is suggested by prior literature, resulting in appraisals of a female subordinate's potential as higher than a male's, advantaging females. Our research design and supplemental analyses help rule out several alternative explanations for these results. We demonstrate that the social context within which appraisals of potential take place can prompt supervisors to differentially use the same accounting information when judging subordinates of different genders.

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

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

I. Introduction

Prior research has long recognized that supervisors use management accounting information to systematically form and incorporate beliefs about what drove subordinates' past performance into appraisals of subordinates' potential to succeed in their current positions (e.g., Shields, Birnberg, & Frieze, 1981). Recent surveys indicate that today, anywhere from 66 percent to 88 percent of large companies also have "high-potential" initiatives in which supervisors appraise subordinates' potential to succeed at different, higher-level positions and use these judgments as a basis for allocating rewards such as training, mentoring, and promotions (Aon Hewitt, 2013; Cappelli & Keller, 2014; Chamorro-Premuzic, Adler, & Kaiser, 2017; Fernández-Aráoz, Roscoe, & Aramaki, 2017; Finkelstein, Costanza, & Goodwin, 2018). Further, prior research largely ignores the social context within which appraisals take place. In today's society and in many businesses, the focus on diversity and inclusion to reduce the underrepresentation of certain groups, such as females in male-dominated fields, can create implicit pressure on supervisors either to recognize or to avoid appearing biased against employees from those groups (Catalyst, 2018; Crandall, Eshleman, and O'Brien, 2002; Krentz, 2019; Monnery & Blais, 2017; Scarborough, 2018). Thus, we investigate whether supervisors' use of past performance information in appraisals of subordinates' potential to succeed at different, higher-level positions differs from what would be expected given prior literature on appraisals for current positions, and whether that differs for subordinates from different groups.

While "potential" has various meanings in practice, we define potential as the probability of a subordinate's success in a future position with more responsibilities and that is at a higher level in the organization (Aon Hewitt, 2013; Silzer & Church, 2009). Because many businesses

are focusing on the underrepresentation of women in higher ranks of the organization, we focus on a context in which the current and higher-level positions are stereotypically “male-typed” (Cejka & Eagly, 1999; Heilman & Caleo, 2018; Johnson, Murphy, Zewdie, & Reichard, 2008). Further, we assume subordinates’ past performance is strong since weak-performing subordinates, regardless of gender, are unlikely to be judged to have high potential.

We first examine whether supervisors differentially interpret identical management accounting information about past performance for male and female subordinates. We then examine whether judgments of potential and reliance on past performance information when making those judgments differ for male and female subordinates. We expect the interpretation of and reliance on past performance information to differ for male and female subordinates because of two features of this context—the subjectivity inherent in appraisals of potential, and the salience of female subordinates’ gender in stereotypically male-typed contexts.

With respect to subjectivity in 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, analyzing, and aggregating information for appraisals (see, e.g., Baker, Jensen, & Murphy, 1988; Bol, 2008, 2011; Bol & Leiby, 2018; DeNisi, Cafferty, & Meglino, 1984; Moers, 2005; Prendergast & Topel, 1993). Relative to appraisals of how well subordinates performed in their current positions, appraisals of their potential to succeed at different, higher-level positions increases information-processing costs and, as a result, subjectivity (Cappelli & Keller, 2014; Shah & Oppenheimer, 2008). For example, supervisors should assess whether and how existing information captures the abilities required for the higher-level position, but in reality, this type of information is hard to quantify and seldom available. At the same time, supervisors should ignore information that does not

capture those abilities, but ignoring this information is notoriously difficult to do (e.g., Benson, Li, & Shue, 2018; Cappelli & Keller, 2014; DeNisi et al., 1984; MacRae & Furnham, 2014; Nisbett, Zukier, & Lemley, 1981; Tetlock, Lerner, & Boettger, 1996).

With respect to the salience of subordinate gender, models of appraisal processes suggest that as relevant information is more difficult to identify and analyze—as it is in appraisals of potential—irrelevant information becomes more salient and more important to judgment processes and outcomes (DeNisi et al., 1984; Ilgen, Barnes-Farrell, & McKellin, 1993; Ilgen & Feldman, 1983). In our context, this is problematic for three reasons. First, when one gender is underrepresented, a subordinate’s gender is particularly difficult to ignore. Second, both men and women stereotypically associate certain types of jobs with males or with females and tend to assume that those of the opposite gender are less capable of performing such jobs (e.g., Biernat & Fuegen, 2001; Cejka & Eagly, 1999; Heilman, 1983, 2001; Heilman & Caleo, 2018). Third, stereotypes are activated with relatively little effort, while discounting them is difficult (Banaji & Greenwald, 1995; Bertrand, Chugh, & Mullainathan, 2005; DeNisi et al., 1984; Foschi, 2000; Ilgen & Feldman, 1983). As a result, in a male-typed context, a female, let alone a strong-performing one, will be especially unexpected, making gender even more salient and reliance on gender or gender stereotypes to simplify the appraisal task more likely (Heilman, 1983; Ilgen & Feldman, 1983; Zarate & Smith, 1990).¹

We rely on attribution theory and socialization theory to develop hypotheses about how a subordinate’s gender influences supervisors’ use of past performance information in appraisals of potential in countervailing ways. First, we focus on supervisors’ interpretation of information

¹ Stereotypes are schemas that include generalizations about groups, and gender stereotypes are generalizations about attributes and roles of females and males. Schemas are used to organize and process information (DiMaggio, 1997).

about a subordinate's past performance in his or her current position. Past performance information is relevant to forecasting a subordinate's potential to succeed in a higher-level position (MacRae & Furnham, 2014; Silzer & Church, 2009). However, since past performance information is inherently noisy, supervisors make inferences (i.e., attributions) about what drove that performance (Birnberg, Frieze, & Shields, 1977; Mitchell & Wood, 1980; Shields et al., 1981; Weiner et al., 1971). These attributions can be impacted by the subordinate's gender. Because individuals tend to stereotypically associate certain jobs with particular genders and expect that those of the opposite gender lack the ability to perform well on those jobs, a female's strong performance in male-typed jobs tends to be attributed to factors other than ability, such as luck (Deaux & Emswiler, 1974; Green & Mitchell, 1979; Greenhaus & Parasuraman, 1993; Lyness & Heilman, 2006). Consequently, we predict that in male-typed contexts, supervisors' attributions of strong performance to ability are lower for female rather than male subordinates.

Second, we focus on supervisors' judgments of a subordinate's potential to succeed in a different, higher-level position and how supervisors incorporate their inferences about the drivers of past performance into those judgments. Prior research demonstrates that judgments of an individual's potential to sustain performance in his or her current position are systematically related to attributions of past performance, such that when attributions to ability are higher, judgments of potential are higher (Kahneman & Frederick, 2002; Shah & Oppenheimer, 2008; Tversky & Kahneman, 1974). This implies that in contexts such as ours, judgments of potential for success in a different, higher-level position would be lower for female than male subordinates, since attributions to ability are lower (e.g., Eagly, Karau, & Makhijani, 1995; Heilman, Block, Martell, & Simon, 1989). However, other research suggests that the social context within which appraisals take place can cause supervisors to perceive that they should

adjust their judgments to conform with organizational norms, strategies, or demands (e.g., Cappelli & Conyon, 2018; Heilman & Caleo, 2018; Levy & Williams, 2004; Lowe, Reckers, & Sanders, 2001). In our setting, this may occur because the contemporary focus on diversity and inclusion has created implicit pressures to avoid appearing biased against, and, in some cases, to place a premium on, members of underrepresented groups (Crandall, et al. 2002; Krentz, 2019; McKinsey, 2017; PwC, 2019). Based on this, we predict that in male-typed contexts, supervisors will rely less on their attributions of past performance when appraising a female subordinate and will judge a strong-performing female subordinate's potential to be higher than a comparably performing male's.

To test our predictions, we conducted a between-subjects experiment in which participants with an average of 20 years of professional work experience assumed the role of a supervisor in the banking industry, a stereotypically male-typed context (Alden, 2014; Bigelow, Lundmark, McLean Parks, & Wuebker, 2014; Catalyst, 2018; Heilman & Caleo, 2018; Holman, Keller, & Colby, 2018; Jaekel & St-Onge, 2016; Zillman, 2019). After reading descriptions of the responsibilities of positions at various ranks in the bank, participants reviewed the past year's performance information for one subordinate. We manipulated whether the metrics and ratings indicated that performance was strong or weak.² We also manipulated the subordinate's gender by either providing no information about the subordinate's gender or by including information that subtly communicated that the subordinate was a male or a female. Participants then judged the subordinate's potential to succeed at the next-higher position in the organization, made causal inferences about the drivers of the subordinate's past performance, and provided their beliefs

² We use only conditions in which the subordinate's current job performance is strong to test our hypotheses. While individuals who are weak performers are unlikely to be viewed as having high potential, we use conditions in which performance is weak in supplemental analyses.

about the current and higher-level positions. We also collect data for one additional condition in which we do not ask participants to appraise a subordinate but ask for the same beliefs about the context and positions that we collect in the other conditions; this provides gender-neutral, appraisal-free beliefs.

Results were consistent with our predictions. First, participants' attributions of a subordinate's strong past performance to ability were significantly lower for female than male subordinates. Second, participants' judgments of potential were higher for strong performers who are female rather than male, and attributions of past performance to ability had a lower association with judgments of potential for female than male subordinates. Further, results in the gender-neutral condition were not different from those in the condition with a male subordinate, validating that participants indeed viewed the context as male-typed. Additional analyses demonstrated that results were not attributable to differential beliefs about the overlap in the skills required for the current and higher-level positions, or about the predictive power of past performance for future potential. Finally, additional data from male, female, and gender-neutral conditions in which the subordinate was a weak performer replicate the pattern of gender-stereotyped attributions of past performance in prior literature and demonstrate that judgments of potential again do not follow from those attributions in a systematic way.

Our study makes several contributions to management accounting research and practice. First, prior accounting research on performance appraisals has largely focused on measurement challenges and on appraisals of the potential to succeed in a current position (see, e.g., Banker & Datar, 1989; Demeré, Sedatole, & Woods, 2019; Feltham & Xie, 1994; Lambert, 2001; Luft, Shields, & Thomas, 2016). However, other research recognizes that there are important differences between these appraisals and those of subordinates' potential to be successful in a

different, higher-level position, especially the typical dearth of information available for the latter. We add to this growing body of accounting literature by showing that implicit pressures may cause supervisors to differentially weight the information that is available to them (Bol & Leiby, 2018; Campbell, 2008; Chan, 2018; Grabner & Moers, 2013).

Second, prior accounting research on appraisals largely ignores subordinate gender and the associated social contexts (exceptions are Bloomfield, Rennekamp, Steenhoven, & Stewart, 2018; Lowe et al., 2001; Maas & Torres-González, 2011). We provide evidence that these factors influence supervisors' use of past performance information in ways that differ from patterns predicted by models of appraisals that ignore gender and social context. Importantly, this research demonstrates a way in which the use of accounting information in organizations can be influenced by broader forces in the organization itself.

Third, our results provide insights that organizations can consider when designing appraisal processes. Specifically, implicit pressures to place a premium on or avoid appearing biased against underrepresented groups may benefit members of such groups in terms of higher appraisal outcomes, but they may do little to mitigate any underlying stereotypes about members of those groups. If supervisors' processing of accounting information and judgments of potential differ for equally performing subordinates of different genders, interventions can be built into the appraisal process that make biases toward one gender, be they negative or positive, more salient so supervisors can try to correct for their effects.

II. Background and Hypothesis Development

Subjectivity in Performance Appraisals

In an appraisal, a supervisor reviews information about a subordinate's performance and then uses that to provide feedback to the subordinate or as an input into decisions about how to

allocate resources among subordinates (Cappelli & Conyon, 2018; DeNisi et al., 1984). Early models of performance appraisal processes focused on the challenges inherent in measuring subordinate performance, then broadened to include supervisors' and subordinates' cognitive processing and, less often, the contexts in which appraisals take place (e.g., their purpose or structure). Given the importance of appraisals to the welfare of both employees and organizations, the goal of much of this research has been to explain why appraisals may be faulty and to identify ways to fix them (Cappelli & Conyon, 2018; DeNisi et al., 1984; Ilgen et al., 1993; Ilgen & Feldman, 1983; Levy & Williams, 2004; Murphy & Cleveland, 1995).

Appraisals of how well a subordinate performed in their current position and their potential for sustained performance in that position require supervisors to identify and interpret performance-relevant information and aggregate it into a judgment. Prior research recognizes that due, in part, to measurement challenges inherent in performance information, supervisors apply subjectivity in their appraisals, for better or worse (see, e.g., Baker et al., 1988; Bol, 2008, 2011; Bol, Kramer, & Maas, 2016; Bol & Leiby, 2018; Bol & Smith, 2011; Cappelli & Conyon, 2018; Demeré et al., 2019; DeNisi et al., 1984; Ittner, Larcker, & Meyer, 2003; Luft et al., 2016; Moers, 2005; Prendergast & Topel, 1993).

Appraisals of subordinates' potential to succeed in a different, higher-level position add even more subjectivity because they require additional processing steps. Supervisors should assess the overlap in the abilities required to be successful in the current and higher-level positions, but assessing overlap is challenging since job descriptions or observations of performance rarely provide explicit signals about this overlap (Benson et al., 2018; Cappelli & Keller, 2014; MacRae & Furnham, 2014; Mattone & Xavier, 2012). Supervisors should identify which existing measures are relevant to performance for the overlapping abilities but ignore

those that are not, which is notoriously difficult to do (Nisbett et al., 1981; Tetlock et al., 1996). They should collect and analyze information for the non-overlapping abilities of the subordinate in question, but this type of information is difficult to measure, if it exists at all (Cappelli & Keller, 2014; Church & Rotolo, 2013; DeNisi et al., 1984; Fairburn & Malcomson, 2001; MacRae & Furnham, 2014; Mattone & Xavier, 2012). Given this difficulty, supervisors are likely to take the less-effortful approach of substituting available-but-noisy past performance information for difficult-to-obtain and even noisier information (Chamorro-Premuzic et al., 2017; Fernández-Aráoz et al., 2017).

Why Subordinate Gender May Impact Supervisors' Appraisal Processes

Models of appraisal processes suggest that as relevant information becomes harder to identify and interpret, irrelevant information becomes more salient and more important to judgment processes and outcomes (DeNisi et al., 1984; Ilgen et al., 1993; Ilgen & Feldman, 1983). Subordinate gender is particularly salient in contexts where certain genders are underrepresented or when the subordinate's job is stereotypically associated with a particular gender (Biernat & Fuegen, 2001; Bigelow et al., 2014; Cejka & Eagly, 1999; Heilman, 1983, 2001; Heilman & Caleo, 2018; Johnson et al., 2008; Koch, D'Mello, & Sackett, 2015). Therefore, supervisors will place subordinates in categories, and in stereotypically gendered contexts (in our case, male-typed), categorization will be based on gender (Ilgen et al., 1993; Ilgen & Feldman, 1983; DeNisi et al., 1984). We expect that categorizing subordinates this way will activate easy-to-access and hard-to-discount gender stereotypes, and these in turn influence supervisors' appraisal processes and judgments (Banaji & Greenwald, 1995; Bertrand et al., 2005; DeNisi et al., 1984; Foschi, 2000; Heilman, 1983; Ilgen & Feldman, 1983; Zarate & Smith, 1990).

With this backdrop, the next sections develop our hypotheses on how gender and gender stereotypes impact supervisors' use of past performance information in appraisals of potential in countervailing ways.

Impact of Gender Stereotypes on Interpretations of Past Performance Information

Because subordinates' past performance information rarely provides precise signals of their contributions to firm success (see, e.g., Banker & Datar, 1989; Feltham & Xie, 1994; Lambert, 2001; Prendergast & Topel, 1993), supervisors must subjectively discern what drove a subordinate's past performance. Prior research finds that supervisors typically attribute performance to four factors—the subordinate's ability and effort, good or bad luck, and the difficulty of the subordinate's job. Ability and effort relate to characteristics of the subordinate (internal attributions), while luck and job difficulty relate to situational factors (external attributions). These factors also differ in terms of their stability, or likelihood that the factor will change in the future. Ability and job difficulty are viewed as being more stable than effort and luck (Birnberg et al., 1977; DeNisi et al., 1984; Harrison, West, & Reneau, 1988; Heider, 1958; Kelley, 1973; Mitchell & Wood, 1980; Shields et al., 1981; Weiner et al., 1971; Weiner & Kukla, 1970).

Causal attributions of past performance can be affected by gender stereotypes, such that in male-typed contexts, individuals expect that females are less capable of strong performance than males (Biernat & Fuegen, 2001; Bigelow et al., 2014; Cejka & Eagly, 1999; Heilman, 1983, 2001; Heilman & Caleo, 2018; Johnson et al., 2008; Koch et al., 2015). Heilman (1983) asserts that if a female's performance is inconsistent with stereotypic expectations, individuals' attributions are biased in ways that allow them to maintain their stereotypic beliefs. As a result, individuals are less likely to attribute performance that is inconsistent with stereotypic

expectations to ability, an internal and stable cause, and more likely to attribute it to external or unstable causes.

Consistent with Heilman (1983), a significant stream of research finds that individuals tend to attribute performance in ways that allow them to maintain their stereotypic beliefs (Deaux & Emswiller, 1974; Feldman-Summers & Kiesler, 1974; Green & Mitchell, 1979; Greenhaus & Parasuraman, 1993). Thus, when an individual's performance is consistent with stereotypic expectations—in our context, when a male performs strongly in a male-typed position—evaluators are more likely to attribute that performance to ability. Conversely, when an individual's performance is inconsistent with stereotypic expectations—in our context, when a female performs strongly in a male-typed position—evaluators are less likely to attribute that performance to ability. Furthermore, for male-typed positions, it takes more evidence to convince individuals that a female's strong performance is due to ability (Brewer, 1988; Deaux & Emswiller, 1974; Feldman-Summers & Kiesler, 1974; Fiske & Neuberg, 1990; Foschi, 2000; Green & Mitchell, 1979; Greenhaus & Parasuraman, 1993; Heilman, 1983; Lyness & Heilman, 2006).

While much of the research above was conducted over two decade ago, we expect to replicate prior results. Stereotypes are notoriously difficult to eradicate in part because individuals are often unaware that they hold them, and stereotypes often influence impressions and judgments outside of conscious awareness (Banaji & Greenwald, 1995; Bertrand et al., 2005). In sum, for female subordinates, we expect a negative impact on attributions to ability when supervisors consider gender in their appraisals.

H1: In male-typed contexts, the extent to which supervisors attribute strong past performance to ability will be lower when the subordinate's gender is female rather than male.

Impact of Gender on Use of Past Performance Information in Judgments of Potential

Prior research finds that attributions of what drove an individual's performance systematically influence evaluators' subsequent judgments about and behaviors toward others (Green & Mitchell, 1979; Harvey, Madison, Martinko, Crook, & Crook, 2014; Heilman & Guzzo, 1978; Kaplan & Reckers, 1985; Pazy, 1987; Pence, Pendleton, Dobbins, & Sgro, 1982; Valle & Frieze, 1976; Weiner et al., 1971). In accounting, Birnberg et al. (1977) and Shields et al. (1981) incorporate this idea into a model of supervisors' use of past performance information in appraisal judgments. They predict and find a positive association between appraisal judgments and attributions to ability. These studies focus on appraisals of a subordinate's potential to sustain performance in his or her current position.

Prior research suggests this model may also be descriptive of appraisals of a subordinate's potential to be successful in a different, higher-level position.³ Recall that relative to appraisals of the potential for sustained performance in a current position, appraisals of potential in a different position are inherently more challenging. When faced with a challenging question, individuals frequently resort to answering an easier but related question (Kahneman & Frederick, 2002; Shah & Oppenheimer, 2008; Tversky & Kahneman, 1974). This suggests that when answering the question about the likelihood a subordinate will be successful in a higher-level position, supervisors will instead answer the related but easier question of the likelihood of continued success in the current position. Thus, one might expect that in our context, judgments of potential would be lower for females than males, consistent with the systematic pattern of the relationship between attributions of past performance and appraisals of potential. Indeed, prior

³ In a recent survey, 75% of managers state that an employees' past performance is *the most* important criteria they consider when appraising the potential for success in different positions, despite knowing that past performance may not predict such future success (Church et al., 2015).

research suggests that females are negatively impacted by judgments resulting from stereotypes (Davidson & Burke, 2000; Eagly & Karau, 2002; Eagly et al., 1995; Heilman et al., 1989; Joshi, Son, & Roh, 2015; Judiesch & Lyness, 1999).

However, this research, as well as that of Birnberg et al. (1977) and Shields et al. (1981), does not incorporate developments in the social context within which appraisals of potential take place. For example, diversity and inclusion, particularly with respect to hiring, retaining, and promoting women in male-dominated fields, has become a “hot topic” in businesses and the business press (Brosnan, 2018; Krentz, 2019; McKinsey, 2017; PwC, 2019; Scarborough, 2018; Shook & Sweet, 2018; Society of Human Resource Management, 2009; Zillman, 2019). For example, a 2017 McKinsey study indicates that 85 percent of companies track gender representation across the levels of their organizations. A 2019 PwC survey notes that 87 percent of the respondents’ organizations have greater diversity and inclusion as an explicit value or priority for their organizations, and prescribes that “...holding leaders accountable for D&I [diversity & inclusion] results...” and “...embedding a diversity lens into talent management...” are essential elements of a cutting-edge focus on diversity and inclusion.

As an emphasis on diversity and inclusion becomes more commonplace, so too does the implicit pressure managers feel to comply with these values or norms (Cappelli & Conyon, 2018; Crandall et al., 2002; Croft & Schmader, 2012; Heilman & Caleo, 2018; Kunda and Spencer, 2003; Levy & Williams, 2004). This can influence supervisors’ judgments of potential in two ways. First, while supervisors may consciously or unconsciously believe females are less able than males to succeed in male-typed positions, supervisors may be lenient toward female subordinates to avoid appearing biased against them (Bear, Cushenbery, London, & Sherman, 2017; Biernat, Tocci, & Williams, 2012; Correll & Simard, 2016). This is more likely to occur

when relevant data is scarce and individuals believe they may have to defend their judgments. In this case, individuals are more likely to selectively use or interpret what information they do have to arrive at judgments consistent with how they believe they should respond (Benson et al., 2018; Tetlock et al., 1996). Supervisors often lack data that allows them to reliably forecast potential regardless of the subordinate's gender (Chamorro-Premuzic et al., 2017; Fernández-Aráoz et al., 2017; Silzer & Church, 2009). However, in male-typed contexts, supervisors may perceive they will be more likely to be called to account for judgments of female than male subordinates, so to mitigate this possibility, they judge females more favorably.

Second, a growing stream of research finds that when a member of an underrepresented group performs well, supervisors can feel pressure to do what they can to retain and reward that individual (Bear et al., 2017; Biernat et al., 2012; Leslie, Manchester, & Dahm, 2014). For example, Leslie et al. (2014) examine a context in which a female performs strongly in a male-dominated organization. They find evidence of a “female premium,” in that females who perform well receive better outcomes (more favorable performance evaluations; more pay) than equally strong-performing males because females are perceived to be more valuable for achieving diversity goals.

Thus, in our context, we expect that social pressures inherent in today's business environment will lead supervisors to judge the potential of female subordinates more favorably than that of equally-strong-performing males.

H2: In male-typed contexts, supervisors' judgments of a subordinate's potential to succeed in a higher-level position will be higher when the subordinate's gender is female rather than male.

We expect that a related effect is that for female subordinates, supervisors will rely less on attributions of past performance to ability than suggested by previous models of appraisal

processes that do not consider gender (Birnberg et al., 1977; Shields et al., 1981). Because of supervisors' conscious or subconscious attempts to avoid appearing biased against females and to indicate they are doing their best to foster diversity and inclusion, they are more concerned with their final judgment of potential rather than on how to systematically aggregate the available information into that judgment. As a result, female subordinates are not negatively impacted by their supervisors' (negative) stereotypical attributions of past performance to ability.

H3: In male-typed contexts, the association between supervisors' judgments of a subordinate's potential to succeed in a higher-level position and attributions of the subordinate's strong performance to ability will be lower when the subordinate's gender is female rather than male.

III. Research Design

Participants

We used business school alumni to proxy for the population of professionals who have experience using performance information to appraise subordinates. With the assistance of a business school's alumni relations office, we recruited alumni to participate in the study. In the summer of 2017, alumni staff emailed a random sample of approximately 4,500 alumni, asking those with two or more years of experience managing employees to participate in a research study; a follow-up email was sent approximately one week later.⁴ The emails included a link to a Qualtrics survey that randomly assigned participants to one of seven experiment conditions (described later) as they clicked on the link. Participants did not receive compensation for participating.

Of the 301 individuals who clicked on the link, 172 participants completed the study. Of those, 160 were representative of the population of interest and completed the task (a response

⁴ The #MeToo Movement began its viral spread in Fall 2017, so we received responses before this feature of the social landscape became even more salient (Johnson & Hawbaker, 2019).

rate of about 3.56 percent of emails sent and 53.16 percent of individuals who clicked on the link).^{5, 6} Consistent with our assumption that this sample was a sound proxy for the population of interest, the 160 participants have an average of 20.31 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 44.03 years, and 25.60 percent are female (Table 1).⁷

[INSERT TABLE 1]

Experiment Design and Procedures

We used a hypothetical case to test our hypotheses.⁸ After providing their informed consent, participants read an overview of a bank's private wealth management group and the hierarchy of positions within it (Appendix A).⁹ They were asked to assume they are a Team Leader within the group. The position descriptions noted that Team Leaders are responsible for ensuring the Portfolio Managers who work for them understand and act consistently with bank strategy, for allocating team resources, and for serving existing and attracting new clients.

We expected participants to perceive the banking context and the positions of Team

⁵ We compare these rates to the same computations using data reported in studies that sought participants with significant work experience, noting that because our solicitation was handled entirely by the alumni association we do not have access to the number of emails that bounced or the number of emails that were opened by recipients (as other studies do). The percentage of usable responses relative to emails sent, inclusive of those that bounced, was 0.7 percent in Anderson and Lillis (2011; more than 10,000 emails sent) and 1.6 percent in Dichev et al. (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 number of emails sent was reported net of those that bounced. In Anderson and Lillis (2011), the percentage of usable responses out of those links opened was 72.17 percent.

⁶ We eliminated twelve 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.

⁷ The percentage of female participants 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 approximately 24% of senior roles (e.g., manager, director, vice president, president; see, e.g., Brosnan, 2018, and Catalyst, 2018).

⁸ The study received Institutional Review Board approval.

⁹ 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 future potential is judged within their organization.

Leader and Portfolio Manager as male-typed, and we include a condition in our design (described later) that allows us to validate that they indeed did so. Careers in finance are stereotypically viewed as male-typed or are dominated by males, and both men and women tend to describe effective finance professionals with attributes that are stereotypically associated with males (Alden, 2014; Bigelow et al., 2014; Catalyst, 2018; Heilman & Caleo, 2018; Holman et al., 2018; Jaekel & St-Onge, 2016; Zillman, 2019).¹⁰ In addition, both men and women tend to associate leaders (in our case, Team Leaders) with stereotypically male traits, which biases against finding results for H2 (Badura, Grijalva, Newman, Yan, & Jeon, 2018; Heilman et al., 1989; Johnson et al., 2008; Koenig, Eagly, Mitchell, & Ristikari, 2011; Offerman & Coats, 2018; Schein, 2007; Schein & Davidson, 1993).

After reading the overview, participants in six of the seven conditions (i.e., excluding those in a control condition who were not asked to appraise a subordinate, as described later) were given the past year's performance information and ratings for one of their Portfolio Managers, which is where we manipulated the gender of the subordinate and the subordinate's past performance (as described with our variables).

The past 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. It also included three qualitative performance notes that could be viewed positively or negatively, providing leeway in interpreting past performance information (although, importantly, far less leeway relative to that available to

¹⁰ Jobs in the banking industry, particularly in wealth management, tend to be male dominated (Zillman, 2019). Females make up less than 50% of banking and financial services employees, with less than 15 percent of executive-level roles held by women (World Economic Forum, 2017).

supervisors in practice, biasing against finding results). Participants were also told that the bank had faced unexpected economic declines in the last three quarters. The latter information was expected to make strong past performance stand out and provided a plausible, external explanation for weak past performance.

Participants then answered a series of questions that were used in analyses (described with our variables) and answered questions about their gender, age, education, and work experience (reported in Table 1). We did not ask explicit questions about beliefs about particular genders, the value of diversity and inclusion, or whether they had felt social pressure to judge females higher than males since responses would likely be biased, particularly since responses, even though anonymous to researchers, were solicited by the alumni association the participants had social ties with.

Independent, Dependent, and Control Variables

We manipulated Portfolio Manager gender in a $(3 \times 2) + 1$ between-subjects design, for a total of seven conditions. Subordinate *Gender* was manipulated at three levels—*Male*, *Female*, and *Neutral*. 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 amongst other task information and questions. 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, so *Neutral* provides a baseline for appraisal judgments and

beliefs absent explicit information about subordinate gender (Appendix B).

We manipulated subordinate *Past Performance* at two levels—*Strong* and *Weak*. *Strong* (*Weak*) performance was operationalized as a majority of “above (below) par” ratings for the five performance measures and an overall “above (below) par” rating (Appendix B). However, across *Strong* and *Weak* conditions, the performance notes were held constant to give participants some leeway in the interpretation of the performance measures and ratings, although far less than is typically available in practice. Since high-potential initiatives in practice and the focus of our hypotheses is a context in which subordinates’ past performance is strong, we use only *Strong* conditions to test our hypotheses. We use *Weak* conditions in supplemental analyses.

Finally, in the “+1” *Control* condition, participants read the same overview of a bank’s private wealth management group and the hierarchy of positions within it, but did not receive past performance information or complete an appraisal for a Portfolio Manager. However, they did answer the same questions about the jobs, context, and demographics as participants in the other six conditions. Thus, *Control* provides a gender-neutral, appraisal-free baseline for beliefs about the current and higher level positions.

We intentionally chose not to manipulate the presence or absence of social pressure. Our theory assumes that in today’s business environment, supervisors face social pressure to appraise females more leniently than males irrespective of the presence of an explicit organizational or external directive. Further, it is virtually impossible to induce feelings of social pressure without making it obvious to participants that we were interested in reactions to subordinate gender, again biasing responses in a socially desirable direction. This choice also biases against finding results for H2 and H3—there is no guarantee that our experienced professional participants have experienced such pressure, or if they did, that they would bring it to bear in our stark laboratory

task.

Our dependent and control variables were captured after the *Gender* and *Past Performance* manipulations. First, participants in the six *Gender* by *Past Performance* conditions (but not *Control*) judged the Portfolio Manager'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?

We used these responses as the dependent variable *Potential* in tests of H2 and H3.

Second, participants in all conditions including *Control* provided their beliefs about two other factors that could impact appraisals of potential. After again reviewing the position descriptions for Portfolio Managers and Team Leaders, they provided their beliefs about the extent of overlap in the skills required to be effective at both (*Overlap*) 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?

They then provided their beliefs about the extent to which past performance is an important predictor of future potential (*PredictivePower*) with this question, anchored on 1 = "Strongly disagree" and 7 = "Strongly agree":

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?"

We used these responses as control variables and/or in supplemental analyses.

Third, in all *Gender* by *Past Performance* conditions, we captured participants' causal attributions of the drivers of the subordinate's past performance by repeating the past performance information and posing this question:

To what extent did the following factors contribute to [Thomas Roan's / Jennifer Roan's / the Portfolio Manager's] above par performance?

Participants allocated 100 points across five causal factors (with more points indicating a more significant driver of past performance): skill/ability, effort, luck, performance target not at appropriate level, and other (with an open-ended response box). We used points allocated to skill/ability (*Ability*) in tests of H1 and H3. We ended with demographic questions.

IV. Results

Hypothesis 1

H1 predicts that in male-typed contexts, the extent to which supervisors attribute strong past performance to ability will be lower when the subordinate's gender is female rather than male. We tested this prediction using the three *Gender* conditions for *Strong Past Performance*. Besides comparing attributions in *Male* and *Female*, we used the *Neutral* condition to validate that participants spontaneously thought of a subordinate as male in this context. Table 2, Panel A provides descriptive statistics for participants' attributions of the subordinate's past performance.

[INSERT TABLE 2]

We tested H1 with an ANOVA and planned comparisons of *Ability*. ANOVA results (Table 2, Panel B) showed a significant effect of *Gender* on *Ability* ($F_{2,65} = 5.52, p = 0.01$).¹¹ Planned comparisons (Table 2, Panel C) showed that *Ability* was significantly lower in *Female* (mean = 36.35) than *Male* (mean = 48.41; $t = -2.70, p = 0.01$, two-tailed). Hence, H1 is supported; female subordinates are disadvantaged relative to males in supervisors' attributions of the drivers of past performance.

In addition, consistent with the assumption that participants perceived our context as male-typed, planned comparisons showed no significant difference in *Ability* in *Male* (mean =

¹¹ For our tests of H1 and H2, all ANOVA assumptions are met.

48.41) and *Neutral* (mean = 49.52; $t = -0.23$, $p = 0.82$, two-tailed), but *Ability* was significantly lower in *Female* (mean = 36.35) than *Neutral* (mean = 49.52; $t = -3.28$, $p < 0.01$, two-tailed).^{12, 13}

Hypothesis 2

H2 predicts that in male-typed contexts, supervisors will judge a subordinate's potential to succeed in a higher-level position as higher for female than male subordinates. We again used the three *Gender* conditions for *Strong Past Performance* to test the hypothesis and validate that participants spontaneously thought of a subordinate as male in this context. Descriptive statistics for *Potential* are in Table 3, Panel A.

[INSERT TABLE 3]

We tested H2 with an ANOVA and planned comparisons of *Potential*. ANOVA results (Table 3, Panel B) showed 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; $t = 2.65$, $p = 0.01$, two-tailed). Hence, H2 is supported; strong-performing female subordinates are advantaged relative to strong-performing males.

In addition, consistent with the assumption that participants perceived our context as male-typed, planned comparisons showed no significant difference in *Potential* in *Male* (mean = 4.41) and *Neutral* (mean = 4.57; $t = -0.34$, $p = 0.74$, two-tailed), but *Potential* was significantly higher in *Female* (mean = 5.35) than *Neutral* (mean = 4.57; $t = 2.05$, $p = 0.05$, two-tailed).

¹² In addition, attributions to “effort,” “luck,” and “unreasonable target” did not differ in the *Male* and *Neutral* conditions (all $p > 0.56$).

¹³ For tests of H1 through H3, statistical inferences remain the same when we control for participants' age. In addition, consistent with prior research on gender stereotypes (Heilman, 2012), statistical inferences remain the same when we control for participants' gender, suggesting that both males and females hold similar stereotypes. Alternately, given that each of our conditions has less than 10 female participants (which is not surprising given statistics about the representation of women in leadership positions), failing to find an effect of participant gender may also be due to a lack of power.

Hypothesis 3

H3 predicts that the association between supervisors' judgments of a subordinate's potential to succeed in a higher-level position and attributions of performance to ability will be lower when the subordinate is female rather than male. We tested H3 with the following regression model using only the *Male* (coded 0) and *Female* (coded 1) *Gender* conditions (not the *Neutral* condition) for *Strong Past Performance*:

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

[INSERT TABLE 4]

We included *Overlap* as a control in the model since—holding past performance constant—the relationship between attributions of performance to ability and the potential to succeed in a higher-level position should take into account beliefs about the overlap in the skills required for both positions. Results are in Table 4. The *Ability* × *Gender* interaction was negative and significant ($\beta_3 = -0.05$, $p = 0.03$, two-tailed), indicating that attributions to ability have a lower association with judgments of potential for female rather than male subordinates. Hence, H3 is supported.

When model (1) was run separately for each of the *Gender* conditions, results (untabulated) showed that *Ability* was a significant predictor of *Potential* in both the *Male* and *Neutral* conditions ($p = 0.01$, two-tailed, for both models), but not in the *Female* condition ($p = 0.72$, two-tailed). Hence, participants did not rely on attributions to ability in their judgments of female subordinates' potential. While, as predicted by prior literature, supervisors' appraisals of a male subordinate's potential are systematically influenced by their beliefs regarding the subordinate's ability, this is not the case when supervisors appraise a female subordinate.

Supplemental Analysis – Other Appraisal-Related Beliefs

To further examine drivers of participants' appraisals, we analyzed their beliefs about the overlap in the skills required for the current and higher-level positions (*Overlap*) and about the predictive power of past performance for future potential (*PredictivePower*). Recall that we asked these questions in all *Gender* and *Past Performance* conditions and in the *Control* condition. In the latter, participants did not receive a subordinate's past performance information nor judge potential, but they were given descriptions of the current and higher-level positions. As a result, beliefs in *Control* were not biased by subordinate gender or the appraisal task. Descriptive statistics for *Overlap* and *PredictivePower* for all three *Gender* conditions for *Strong Past Performance* and for the *Control* condition are in Table 5, Panel A.

[INSERT TABLE 5]

With respect to beliefs about the overlap in the skills required for the current and higher-level positions, an ANOVA (Table 5, Panel B) found no significant differences in *Overlap* across the *Gender* and *Control* conditions ($F_{3,90} = 0.26, p = 0.85$). We also noted that when *Overlap* was included in the model of the relationships between judgments of potential, attributions to ability, and subordinate gender for testing H3, *Overlap* was not significant ($p = 0.57$, two-tailed; Table 4). Finally, we reran the model used to test H3 with an *Overlap* \times *Gender* interaction; the interaction was not significant ($p = 0.55$, two-tailed; not tabulated). In sum, participants did not incorporate beliefs about the overlap in skills required for success in the current and higher-level positions in their judgments of potential, nor did they believe that the skills overlap was different for female and male subordinates.

It is possible that our results were driven by the fact that participants stereotypically believe that females have certain attributes that are important for success in the higher-level

position that are not captured by past performance (e.g., communal; compassionate), and thus that the predictive power of past performance is lower. We performed analyses to rule out this alternative explanation using participants' beliefs about the predictive power of past performance for future potential.

An ANOVA (Table 5, Panel C) found no significant differences in *PredictivePower* across the *Gender* and *Control* conditions ($F_{3,90} = 0.93, p = 0.43$). In addition, we reran the regression model used to test H3, but substituted *PredictivePower* for *Ability*:

$$(2) \text{ Potential} = \beta_0 + \beta_1(\text{PredictivePower}) + \beta_2(\text{Gender}) + \beta_3(\text{PredictivePower} \times \text{Gender}) + \beta_4(\text{Overlap}) + \varepsilon$$

The *PredictivePower* \times *Gender* interaction was not significant ($p = 0.49$, two-tailed; not tabulated). We also reran the model used to test H3 with *PredictivePower* and a *PredictivePower* \times *Gender* interaction. *PredictivePower* and the *PredictivePower* \times *Gender* interaction were not significant (both $p > 0.35$, two-tailed; not tabulated), but the *Ability* \times *Gender* interaction remained negative and significant ($p < 0.01$). As a result, differential beliefs about the extent to which past performance is a useful predictor of potential for females and males were not driving our results.

Supplemental Analysis – Attributions and Judgments of Potential for Weak Performers

Recall that while we expect supervisors to stereotypically believe females are less able than males to succeed in male-typed positions, we also expect that supervisors will be more lenient to female than male subordinates to avoid the appearance of being biased against them (Bear et al., 2017; Biernat et al., 2012; Correll & Simard, 2016).

To examine a potential boundary condition of this expectation, we use the *Weak Past Performance* conditions that were identical to those used in our primary study, except that measures of the subordinate's past performance and associated ratings were held constant as

weak (but the performance notes were the same as in *Strong Past Performance*). While weak performers are unlikely to be viewed as having high potential, these conditions allow us to test the limits of when gender and gender stereotypes affect these appraisals. Table 6, Panel A provides descriptive statistics for our primary dependent measures, *Ability* and *Potential*, in the three *Gender* conditions for *Weak Past Performance*.

[INSERT TABLE 6]

Consistent with the theory used to develop H1, we expect that when a subordinate's past performance is consistent with stereotypic expectations—in our context, when a female performs poorly in a male-typed position—supervisors will attribute that weak performance to a lack of ability to a greater extent for females than for equally weak-performing males. ANOVA results (Table 6, Panel B) showed a significant effect of *Gender* on *Ability* ($F_{2,66} = 7.60, p = 0.01$). Untabulated planned comparisons showed that attributions of weak past performance to ability were higher in *Female* (mean = 42.05) than *Male* (mean = 25.80; $t = 3.33, p = 0.01$, two-tailed). Hence, with respect to attributions of the drivers of past performance, female subordinates are negatively impacted by supervisors' incorporation of gender stereotypes in their appraisals, in that supervisors believe weak past performance is driven by a lack of ability to a greater extent for female than male subordinates.

Consistent with the theory used to develop H2, we expect that supervisors will judge weak-performing female subordinates more leniently than they will equally weak-performing male subordinates, despite the fact that they are believed to be less able. ANOVA results (Table 6, Panel C) showed that *Gender* did not have a significant effect on *Potential* ($F_{2,68} = 0.76, p = 0.47$). Untabulated planned comparisons showed that *Potential* was not significantly different in *Female* (mean = 3.71) than *Male* (mean = 3.48; $t = 0.85, p = 0.40$, two-tailed). Note that, unlike

what was found in our strong-performing conditions, there is no evidence of a “female premium” in the weak-performing conditions, presumably because a poor-performing subordinate, regardless of gender, is unlikely to be perceived as having any special value to the organization. Failing to observe a “female premium” in the weak-performing conditions further suggests that supervisors do not assign females higher appraisals simply to move them out from under their supervision.

Finally, we perform the same regression analyses used to test H3 to provide further evidence that supervisors are more lenient and less systematic when appraising a weak-performing female subordinate’s potential. When performance is weak, we again find that *Ability* was a significant predictor of *Potential* in both the *Male* and *Neutral* conditions ($p = 0.02$ and $p = .01$, two-tailed, untabulated for both models), but not in the *Female* condition ($p = 0.40$, two-tailed, untabulated). This provides further evidence that a desire to avoid appearing biased against females causes supervisors to be more concerned with their final judgments of potential rather than on how to systematically incorporate available information into that judgment.

V. Discussion and Conclusion

Using a stereotypically male-typed context, we examine how supervisors use management accounting information about subordinates’ past performance when appraising their potential to succeed in a different, higher-level position, and whether supervisors’ use of performance information differs depending on a subordinate’s gender. We predict and find that because of the social context within which these appraisals are conducted, gender and gender stereotypes influence appraisal processes in countervailing ways. On the one hand, when supervisors analyze past performance information to subjectively determine what factors drove the subordinate’s past performance, gender stereotypes negatively affect the extent to which

supervisors attribute past performance to ability for females relative to males, disadvantaging females. On the other hand, implicit pressures to avoid appearing biased against certain groups or to place a premium on strong-performing members of those groups to demonstrate a commitment to diversity and inclusion positively affects females relative to males—supervisors judge strong performing female subordinates to have higher potential than equally strong performing male subordinates, advantaging females. Importantly, our results are not driven by differences in participants' beliefs about the predictive power of past performance for female and male subordinates or the overlap in the abilities required for the current and higher-level positions.

Our study has several limitations that provide opportunities for additional research. First, we infer that more favorable appraisal outcomes for female relative to male subordinates were driven by implicit pressures to judge females more favorably in male-typed contexts, either to avoid the appearance of bias or because a strong-performing female is especially valuable in a male-typed context. We do not induce this pressure, nor measure the extent to which participants feel this pressure or their beliefs about the value of diversity and inclusion. Doing so would almost surely yield socially acceptable responses, and, notably, we find results even without these interventions. Future research could attempt to manipulate social pressure or solicit these views from a comparable population that would not have the impression management concerns that our participants likely had. Second, our experimental context is stark, and we only examine one side of our theory (i.e., we do not conduct a parallel experiment with male subordinates in a stereotypically female-typed context). This was intentional—we wanted to be respectful of the time of professional participants, increase control over inferences of how past performance information is used in appraisals, and focus on a common social context in businesses today (the

underrepresentation of women in stereotypically male-typed organizations). However, our task necessarily excludes an array of other factors that supervisors might want to incorporate into their judgments of potential, and we do not ask participants to appraise the potential of multiple subordinates at once, which is likely done in practice. Future research could extend this study to ascertain what information supervisors would like to have or would choose to use when appraising potential, and how our results might differ when there are multiple subordinates to be appraised. Third, we do not ask participants to use their judgments of potential to then make decisions about rewards, nor do we explicitly say that the judgments of potential would be used for this purpose, since doing so would necessarily require a comparison of multiple subordinates and make gender particularly salient. It is possible that such decisions would not follow the patterns we observe. Finally, the order in which we asked task-related questions could have impacted responses. Indeed, asking supervisors to explicitly consider their performance attributions or the overlap in skills required to be successful in the current and higher-level positions may prompt more systematic processing, serving as a debiasing mechanism for overly favorable or unfavorable judgments for a particular gender. Organizations would be well served by research that examines debiasing mechanisms in contexts like ours, as well as research that examines the implications of our results for subordinates' motivation, job satisfaction, and productivity.

Our research contributes to a growing body of accounting literature that focuses on appraisals of subordinates' potential to succeed in a different, higher-level position, rather than their potential to succeed in their current position. It is important to recognize that these are fundamentally different appraisal tasks, and as such, supervisors may use the same accounting information in different ways for each. In addition, few accounting studies on appraisals consider

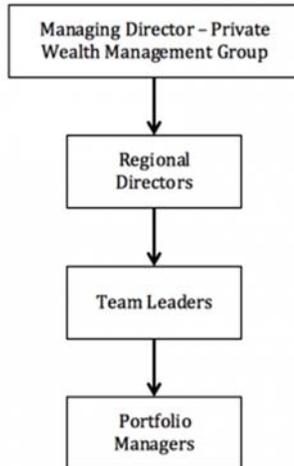
the impact of subordinate gender on appraisal processes and outcomes, but understanding whether and how the salience of gender in today's social environment affects the use of accounting information sheds light on how we can design more effective performance measurement and control systems.

Finally, our results suggest that efforts to increase the representation of females in progressively higher-level positions in certain organizations may make it more likely that strong-performing females will be identified as having high potential, but may do little in the way of eliminating underlying stereotypes about their ability. If firms do use these appraisals to allocate more resources such as training, mentoring, and promotions to subordinates who are judged to have high potential, then females may benefit from (and males may be harmed by) these efforts. However, if, as suggested by anecdotal evidence, mentoring and training opportunities are instead allocated on the basis of beliefs about subordinates' ability, females who are labelled as having high potential may still receive fewer of these opportunities than similarly identified males, making it harder for them to obtain the skills necessary for promotion. For example, Ibarra, Carter, and Silva (2010) find that while high-potential females are just as likely as their male counterparts to have mentors, females tend to be paired with mentors who have less organization clout and are less likely to actively advocate on their mentees' behalf. As such, we provide insights that can help organizations consider how to design appraisal processes to neutralize any gender biases, positive or negative, that supervisors may bring to the task.

APPENDIX A

Overview from Experiment Instrument

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

Gender and Past Performance Manipulations from Experiment Instrument

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

Gender = Female

Gender = Neutral

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



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



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	PERFORMANCE TARGET	INDIVIDUAL PERFORMANCE RATING:		
			BELOW PAR	PAR	ABOVE PAR
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
<i>NOTES:</i> Takes time to adapt to new systems and processes; can be easily flustered. Very direct and straightforward. Seeks advice from others when it is appropriate to do so.					
OVERALL PERFORMANCE RATING					X

Past Performance = Weak

	INDIVIDUAL PERFORMANCE	PERFORMANCE TARGET	INDIVIDUAL PERFORMANCE RATING:		
			BELOW PAR	PAR	ABOVE PAR
Number of New Clients	4	6	X		
Number of Lost Clients	0	1			X
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		
<i>NOTES:</i> Takes time to adapt to new systems and processes; can be easily flustered. Very direct and straightforward. Seeks advice from others when it is appropriate to do so.					
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.

References

- Alden, W. (2014, September 30). Wall Street's young bankers are still mostly white and male, report says. *The New York Times*. Retrieved from <https://dealbook.nytimes.com/2014/09/30/wall-streets-young-bankers-are-still-mostly-white-and-male/>
- Anderson, S. W., & Lillis, A. M. (2011). Corporate frugality: Theory, measurement and practice. *Contemporary Accounting Research*, 28, 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, UK: Aon Hewitt.
- Badura, K. L., Grijalva, E., Newman, D. A., Yan, T. T., & Jeon, G. (2018). Gender and leadership emergence: A meta-analysis and explanatory model. *Personnel Psychology*, 71, 335–367. <https://doi.org/10.1111/peps.12266>
- Baker, G. P., Jensen, M. C., & Murphy, K. J. (1988). Compensation and incentives: Practice vs. theory. *The Journal of Finance*, 43, 593–616. <https://doi.org/10.1111/j.1540-6261.1988.tb04593.x>
- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, 68, 181–198. <https://doi.org/10.1037/0022-3514.68.2.181>
- Banker, R. D., & Datar, S. M. (1989). Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research*, 27, 21–39. <https://doi.org/10.2307/2491205>
- Bear, J. B., Cushenbery, L., London, M., & Sherman, G. D. (2017). Performance feedback, power retention, and the gender gap in leadership. *The Leadership Quarterly*, 28, 721–740. <https://doi.org/10.1016/j.leaqua.2017.02.003>
- Benson, A., Li, D., & Shue, K. (2018). *Promotions and the Peter Principle* (National Bureau of Economic Research Working Paper Series, paper 24343). Cambridge, MA: NEBR.
- Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95(2), 94–98. <https://doi.org/10.1257/000282805774670365>
- Biernat, M., & Fiegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues*, 57, 707–724. <https://doi.org/10.1111/0022-4537.00237>
- Biernat, M., Tocci, M. J., & Williams, J. C. (2012). The language of performance evaluations: Gender-based shifts in content and consistency of judgment. *Social Psychological and Personality Science*, 3, 186–192. <https://doi.org/10.1177/1948550611415693>
- Bigelow, L., Lundmark, L., McLean Parks, J., & Wuebker, R. (2014). Skirting the issues: Experimental evidence of gender bias in IPO prospectus evaluations. *Journal of Management*, 40, 1732–1759. <https://doi.org/10.1177/0149206312441624>
- Birnberg, J. G., Frieze, I. H., & Shields, M. D. (1977). The role of attribution theory in control systems. *Accounting, Organizations and Society*, 2, 189–200. [33](https://doi.org/10.1016/0361-</p></div><div data-bbox=)

[3682\(77\)90011-3](#)

- Bloomfield, R., Rennekamp, K., Steenhoven, B., & Stewart, S. (2018). *Gender bias through recategorization of financial analysts* (Working paper). Ithaca, NY: Cornell University.
- Bol, J. (2008). Subjectivity in compensation contracting. *Journal of Accounting Literature*, 27, 1–32.
- Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86, 1549–1575. <https://doi.org/10.2308/accr-10099>
- Bol, J. C., Kramer, S., & Maas, V. S. (2016). How control system design affects performance evaluation compression: The role of information accuracy and outcome transparency. *Accounting, Organizations and Society*, 51, 64–73. <https://doi.org/10.1016/j.aos.2016.01.001>
- Bol, J. C., & Leiby, J. (2018). Subjectivity in professionals' incentive systems: Differences between promotion- and performance-based assessments. *Contemporary Accounting Research*, 35, 31–57. <https://doi.org/10.1111/1911-3846.12393>
- Bol, J. C., & Smith, S. D. (2011). Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review*, 86, 1213–1230. <https://doi.org/10.2308/accr-10038>
- Brewer, M. (1988). A dual process model of impression formation. In T. Srull and R. Wyer (Eds.), *Advances in social cognition* (pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum.
- Brosnan, A. (2018). *Women in business: Beyond policy to progress*. Retrieved from Grant Thornton International Ltd. website: <https://www.grantthornton.global/globalassets/1.-member-firms/global/insights/women-in-business/grant-thornton-women-in-business-2018-report.pdf>
- Campbell, D. (2008). Nonfinancial performance measures and promotion-based incentives. *Journal of Accounting Research*, 46, 297–332. <https://doi.org/10.1111/j.1475-679x.2008.00275.x>
- Cappelli, P., & Conyon, M. (2018). What do performance appraisals do? *ILR Review*, 71, 88–116. <https://doi.org/10.1177/0019793917698649>
- Cappelli, P., & Keller, J. (2014). Talent management: Conceptual approaches and practical challenges. *Annual Review of Organizational Psychology and Organizational Behavior*, 1, 305–331. <https://doi.org/10.1146/annurev-orgpsych-031413-091314>
- Catalyst. (2018). *Quick take: Women in management*. Retrieved March 20, 2019 from Catalyst website: <https://www.catalyst.org/research/women-in-management/>
- Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, 25, 413–423. <https://doi.org/10.1177/0146167299025004002>
- Chamorro-Premuzic, T., Adler, S., & Kaiser, R. (2017, October 3). What science says about identifying high-potential employees. *Harvard Business Review*. Retrieved from <https://hbr.org/2017/10/what-science-says-about-identifying-high-potential-employees>
- Chan, E. W. (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. H., & Rotolo, C. T. (2013). How are top companies assessing their high-potentials and senior executives? A talent management benchmark study. *Consulting Psychology Journal: Practice and Research*, 65, 199–223. <https://doi.org/10.1037/a0034381>
- Correll, S., & Simard, C. (2016, April 29). Vague feedback is holding women back. *Harvard Business Review*. Retrieved from <https://hbr.org/2016/04/research-vague-feedback-is-holding-women-back>
- Crandall, C.S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82, 359-378. <https://doi.org/10.1037/0022-3514.82.3.359>
- Croft, A., & Schmader, T. (2012). The feedback withholding bias: Minority students do not receive critical feedback from evaluators concerned about appearing racist. *Journal of Experimental Social Psychology*, 48, 1139-1144. <https://doi.org/10.1016/j.jesp.2006.06.007>.
- Davidson, M., & Burke, R. (2000). *Women in management: Current research issues (Volume II)*. London, UK: Sage Publications Ltd.
- Deaux, K., & Emswiller, T. (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, 80–85. <https://doi.org/10.1037/h0035733>
- Deméré, B. W., Sedatole, K. L., & Woods, A. (2019). The role of calibration committees in subjective performance evaluation systems. *Management Science*, 65, 1562–1585. <https://doi.org/10.1287/mnsc.2017.3025>
- DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance*, 33, 360–396. [https://doi.org/10.1016/0030-5073\(84\)90029-1](https://doi.org/10.1016/0030-5073(84)90029-1)
- Dichev, I. D., Graham, J. R., Harvey, C. R., & Rajgopal, S. (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>
- DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology*, 23, 263–287. <https://doi.org/10.1146/annurev.soc.23.1.263>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573–598. <https://doi.org/10.1037/0033-295x.109.3.573>
- Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, 117, 125–145. <https://doi.org/10.1037/0033-2909.117.1.125>
- Fairburn, J. A., & Malcomson, J. M. (2001). Performance, promotion, and the Peter Principle. *The Review of Economic Studies*, 68, 45–66. <https://doi.org/10.1111/1467-937x.00159>
- Feldman-Summers, S., & Kiesler, S. B. (1974). Those who are number two try harder: The effect of sex on attributions of causality. *Journal of Personality and Social Psychology*, 30, 846–855. <https://doi.org/10.1037/h0037604>
- Feltham, G., & Xie, J. (1994). Performance measure congruity and diversity in multi-task

- principal/agent relations. *The Accounting Review*, 69(3): 429–453.
- Fernández-Aráoz, C., Roscoe, A., & Aramaki, K. (2017). Turning potential into success: The missing link in leadership development. *Harvard Business Review*, 95(6), 86–93.
- Finkelstein, L. M., Costanza, D. P., & Goodwin, G. F. (2018). Do your high potentials have potential? The impact of individual differences and designation on leader success. *Personnel Psychology*, 71, 3–22. <https://doi.org/10.1111/peps.12225>
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, 23, 1–74.
- Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, 26, 21–42. <https://doi.org/10.1146/annurev.soc.26.1.21>
- Grabner, I., & Moers, F. (2013). Managers' choices of performance measures in promotion decisions: An analysis of alternative job assignments. *Journal of Accounting Research*, 51, 1187–1220. <https://doi.org/10.1111/1475-679x.12027>
- Graham, J. R., Harvey, C. R., & Puri, M. (2013). Managerial attitudes and corporate actions. *Journal of Financial Economics*, 109, 103–121. <https://doi.org/10.1016/j.jfineco.2013.01.010>
- Green, S. G., & Mitchell, T. R. (1979). Attributional processes of leaders in leader–member interactions. *Organizational Behavior and Human Performance*, 23, 429–458. [https://doi.org/10.1016/0030-5073\(79\)90008-4](https://doi.org/10.1016/0030-5073(79)90008-4)
- Greenhaus, J. H., & Parasuraman, S. (1993). Job performance attributions and career advancement prospects: An examination of gender and race effects. *Organizational Behavior and Human Decision Processes*, 55, 273–297. <https://doi.org/10.1006/obhd.1993.1034>
- Harrison, P., West, S., & Reneau, J. (1988). Initial attributions and information-seeking by superiors and subordinates in production variance investigations. *The Accounting Review*, 63, 307–320.
- Harvey, P., Madison, K., Martinko, M., Crook, T. R., & Crook, T. A. (2014). Attribution theory in the organizational sciences: The road traveled and the path ahead. *Academy of Management Perspectives* 28, 128–146. <https://doi.org/10.5465/amp.2012.0175>
- 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. *Research in Organizational Behavior*, 5, 269–298.
- Heilman, M. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57, 657–674.
- Heilman, M. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32, 11–135. <https://doi.org/10.1016/j.riob.2012.11.003>
- Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology*, 74, 935–942. <https://doi.org/10.1037/0021-9010.74.6.935>

- Heilman, M. E., & Caleo, S. (2018). Combatting gender discrimination: A lack of fit framework. *Group Processes & Intergroup Relations*, 21, 725–744. <https://doi.org/10.1177/1368430218761587>
- Heilman, M. E., & Guzzo, R. A. (1978). The perceived cause of work success as a mediator of sex discrimination in organizations. *Organizational Behavior and Human Performance*, 21, 346–357. [https://doi.org/10.1016/0030-5073\(78\)90058-2](https://doi.org/10.1016/0030-5073(78)90058-2)
- Holman, J., Keller, L., & Colby, L. (2018, February 6). Banks aren't showing their math on identical 1% gender pay gaps. *Bloomberg*. Retrieved from <https://www.bloomberg.com/news/articles/2018-02-06/four-wall-street-banks-all-reveal-a-1-pay-gap-coincidence>
- Ibarra, H., Carter, N. M., & Silva, C. (2010). Why men still get more promotions than women. *Harvard Business Review*, 88, 80-85. <https://hbr.org/2010/09/why-men-still-get-more-promotions-than-women>.
- Ilgen, D., Barnes-Farrell, J. L., & McKellin, D. B. (1993). Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes*, 54, 321–368. <https://doi.org/10.1006/obhd.1993.1015>
- Ilgen, D., & Feldman, J. (1983). Performance appraisal: A process focus. *Research in Organizational Behavior*, 5, 141–197.
- Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*, 78, 725–758. <https://doi.org/10.2308/accr.2003.78.3.725>
- Jaekel, A., & St-Onge, E. (2016, October 25). Why women aren't making it to the top of financial services firms. *Harvard Business Review*. Retrieved from <https://hbr.org/2016/10/why-women-arent-making-it-to-the-top-of-financial-services-firms>
- Johnson, C., & Hawbaker, K. (2019, July 10). #MeToo: A timeline of events. *Chicago Tribune*. Retrieved from: <https://www.chicagotribune.com/lifestyles/ct-me-too-timeline-20171208-htmlstory.html>
- Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. (2008). The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes*, 106, 39–60. <https://doi.org/10.1016/j.obhdp.2007.12.002>
- Joshi, A., Son, J., & Roh, H. (2015). When can women close the gap? A meta-analytic test of sex differences in performance and rewards. *Academy of Management Journal*, 58, 1516–1545. <https://doi.org/10.5465/amj.2013.0721>
- Judiesch, M. K., & Lyness, K. S. (1999). Left behind? The impact of leaves of absence on managers' career success. *Academy of Management Journal*, 42, 641–651. <https://doi.org/10.2307/256985>
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY: Cambridge University Press.

- Kaplan, S., & Reckers, P. (1985). An examination of auditor performance evaluation. *The Accounting Review*, 60, 477–487.
- Kelley, H. (1973). The processes of causal attribution. *American Psychologist*, 28, 107–128. <https://doi.org/10.1037/h0034225>
- Koch, A. J., D’Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100, 128–161. <https://doi.org/10.1037/a0036734>
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137, 616–642. <https://doi.org/10.1037/a0023557>
- Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, 129, 522–544. <https://psycnet.apa.org/doi/10.1037/0033-2909.129.4.522>
- Krentz, M. (2019, February 5). Survey: What diversity and inclusion policies do employees actually want? *Harvard Business Review*. Retrieved from <https://hbr.org/2019/02/survey-what-diversity-and-inclusion-policies-do-employees-actually-want>
- Lambert, R. A. (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)
- Leslie, L. M., Manchester, C., & Dahm, P. C. (2014). Promoting discrimination by valuing diversity? A reversal of the gender gap among high potentials. *Academy of Management Proceedings*, 2014, 10992. <https://doi.org/10.5465/ambpp.2014.10992abstract>
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30, 881–905. <https://doi.org/10.1016/j.jm.2004.06.005>
- Lowe, D. J., Reckers, P. M., & Sanders, D. (2001). The influence of gender, ethnicity, and individual differences on perceptions of career progression in public accounting. *International Journal of Auditing*, 5, 53–71. <https://doi.org/10.1111/1099-1123.00325>
- Luft, J., Shields, M. D., & Thomas, T. F. (2016). Additional information in accounting reports: Effects on management decisions and subjective performance evaluations under causal ambiguity. *Contemporary Accounting Research*, 33, 526–550. <https://doi.org/10.1111/1911-3846.12156>
- Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91, 777–785. <https://doi.org/10.1037/0021-9010.91.4.777>
- Maas, V. S., & Torres-González, R. (2011). Subjective performance evaluation and gender discrimination. *Journal of Business Ethics*, 101, 667–681. <https://doi.org/10.1007/s10551-011-0763-7>
- MacRae, I., & Furnham, A. (2014). *High potential: How to spot, manage and develop talented people at work*. London, UK: Bloomsbury Publishing.

- Mattone, J., & Xavier, L. (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.
- Mitchell, T. R., & Wood, R. E. (1980). Supervisor's responses to subordinate poor performance: A test of an attributional model. *Organizational Behavior and Human Performance*, 25, 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, 67–80. <https://doi.org/10.1016/j.aos.2003.11.001>
- Monnery, L., & Blais, L. (2017, January 1). To get diversity right, get potential right. Retrieved from the Egon Zehnder website: <https://www.egonzehnder.com/insight/to-get-diversity-right-get-potential-right>
- Murphy, K., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage Publications, Inc.
- Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, 13, 248–277. [https://doi.org/10.1016/0010-0285\(81\)90010-4](https://doi.org/10.1016/0010-0285(81)90010-4)
- Pazy, A. (1987). Sex differences in responsiveness to organizational career management. *Human Resource Management*, 26, 243–256. <https://doi.org/10.1002/hrm.3930260208>
- Pence, E. C., Pendleton, W. C., Dobbins, G. H., & Sgro, J. A. (1982). Effects of causal explanations and sex variables on recommendations for corrective actions following employee failure. *Organizational Behavior and Human Performance*, 29, 227–240. [https://doi.org/10.1016/0030-5073\(82\)90257-4](https://doi.org/10.1016/0030-5073(82)90257-4)
- Prendergast, C., & Topel, R. (1993). Discretion and bias in performance evaluation. *European Economic Review*, 37, 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 & inclusion survey. *PwC Global*. Retrieved from PwC's website: <https://www.pwc.com/gx/en/services/people-organisation/global-diversity-and-inclusion-survey.html#data>.
- Scarborough, W. (2018, February 23). What the data says about women in management between 1980 and 2010. *Harvard Business Review*. Retrieved from <https://hbr.org/2018/02/what-the-data-says-about-women-in-management-between-1980-and-2010>
- Schein, V. E. (2007). Women in management: Reflections and projections. *Women in Management Review*, 22, 6–18. <https://doi.org/10.1108/09649420710726193>
- Schein, V. E., & Davidson, M. J. (1993). Think manager, think male. *Management Development Review*, 6. <https://doi.org/10.1108/eum0000000000738>
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, 134, 207–222. <https://doi.org/10.1037/0033-2909.134.2.207>
- Shields, M. D., Birnberg, J. G., & Frieze, I. H. (1981). Attributions, cognitive processes and

- control systems. *Accounting, Organizations and Society*, 6, 69–93.
[https://doi.org/10.1016/0361-3682\(81\)90023-4](https://doi.org/10.1016/0361-3682(81)90023-4)
- Shook, E., & Sweet, J. (2018). *When she rises, we all rise – Getting to equal 2018: Creating a culture where everyone thrives*. Dublin, Ireland: Accenture.
- Silzer, R., & Church, A. H. (2009). The pearls and perils of identifying potential. *Industrial and Organizational Psychology* 2(4): 377-412.
- Society for Human Resource Management. (2009). *Global diversity and inclusion: Perceptions, practices and attitudes*. Retrieved from https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/09_Diversity_and_Inclusion_Report.pdf
- Tetlock, P., Lerner, J., & Boettger, R. (1996). The dilution effect: Judgmental bias, conversational convention, or a bit of both? *European Journal of Social Psychology*, 2, 377–412. <https://doi.org/10.1111/j.1754-9434.2009.01163>.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Valle, V. A., & Frieze, I. H. (1976). Stability of causal attributions as a mediator in changing expectations for success. *Journal of Personality and Social Psychology*, 33, 579–587. <https://doi.org/10.1037/0022-3514.33.5.579>
- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. (1971). Perceiving the causes of success and failure. In E. Jones, D. Kanouse, H. Kelley, R. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.
- Weiner, B., & Kukla, A. (1970). An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology*, 15, 1–20. <https://doi.org/10.1037/h0029211>
- Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, 8, 161–185. <https://doi.org/10.1521/soco.1990.8.2.161>
- Zillman, C. (2019, September 19). Wall Street Has Never Had a Woman CEO. Why Not? *Fortune*. Retrieved from <https://fortune.com/longform/banking-finance-women-ceos-wall-street/>

TABLE 1
Participant Demographic Data

	Mean	Std. Deviation
Age, in years	44.03	9.54
Professional work experience, in years	20.31	8.80
Count Percent		
Participant Gender:		
Male	119	74.40%
Female	41	25.60%
<i>Total</i>	<i>160</i>	<i>100.00%</i>
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%
<i>Total</i>	<i>160</i>	<i>100.00%</i>

TABLE 2
Hypothesis 1: Supervisors' Attributions of Strong Past Performance to Ability

Panel A: Mean (Standard Deviation) of Causal Attributions of Past Performance in *Strong Past Performance*^a

Points Allocated to Causal Factor ^b	<i>Gender</i> ^a		
	<i>Male</i> n = 22	<i>Female</i> n = 23	<i>Neutral</i> n = 21
Ability	48.41 (17.21)	36.35 (12.54)	49.52 (14.13)
Effort	34.55 (15.73)	38.13 (13.62)	33.10 (12.99)
Luck	7.73 (7.68)	15.83 (17.70)	8.81 (10.11)
Unreasonable Target	7.95 (10.76)	9.26 (8.33)	7.86 (14.80)
Other	1.36 (4.67)	0.43 (2.09)	0.71 (2.39)

Panel B: ANOVA – Effect of *Gender* on *Ability* in *Strong Past Performance*^a

Source of Variation	df	Mean Square	<i>F</i>	<i>p</i>
<i>Gender</i>	2	1197.24	5.52	0.01
Error	63	217.04		
Total	65			

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

<i>Gender</i>	Difference	Std. Error	<i>t</i>	df	<i>p</i> (two-tailed)
<i>Female v. Male</i>	(12.06)	4.47	(2.70)	43	0.01
<i>Male v. Neutral</i>	(1.11)	4.82	(0.23)	41	0.82
<i>Female v. Neutral</i>	(13.17)	4.02	(3.28)	42	< 0.01

Notes

^a 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*.

^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.

TABLE 3

Hypothesis 2: Influence of Subordinate Gender on Supervisors' Judgments of Potential

Panel A: Mean (Standard Deviation) of *Potential in Strong Past Performance*^a

	<i>Gender</i> ^a		
	<i>Male</i> n = 22	<i>Female</i> n = 23	<i>Neutral</i> n = 21
<i>Potential</i> ^b	4.41 (1.50)	5.35 (0.78)	4.57 (1.63)

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

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

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

<i>Gender</i>	Difference	Std. Error	<i>t</i>	df	<i>p</i> (two-tailed)
<i>Female v. Male</i>	0.94	0.35	2.65	43	0.01
<i>Male v. Neutral</i>	(0.16)	0.48	(0.34)	41	0.74
<i>Female v. Neutral</i>	0.78	0.38	2.05	42	0.05

Notes

^a 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*.

^b We measured participants' judgments 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?"

TABLE 4
Hypothesis 3: Association Between Judgments of Potential and Attributions to Ability

Regression Model with *Male* and *Female* Conditions in *Strong Past Performance* Only ^c

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

Variable	β	p (two-tailed)
Intercept	1.80	0.05
<i>Ability</i> ^b	0.05	< 0.01
<i>Gender</i> ^c	3.36	< 0.01
<i>Ability</i> \times <i>Gender</i>	(0.05)	0.03
<i>Overlap</i> ^d	0.08	0.57
n = 45		
$R^2 = 0.34$		
Adjusted $R^2 = 0.28$		

Notes

- ^a We measured participants' judgments 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 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 measure 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?"

TABLE 5
Supplemental Analysis – Other Appraisal-Related Beliefs

Panel A: Mean (Standard Deviation) of *Overlap* and *PredictivePower* in *Strong Past Performance*^a

	<i>Gender</i> ^a			<i>Control</i> n = 25
	<i>Male</i> n = 22	<i>Female</i> n = 23	<i>Neutral</i> n = 21	
<i>Overlap</i> ^b	4.32 (1.32)	4.04 (0.93)	4.10 (1.04)	4.08 (1.26)
<i>PredictivePower</i> ^c	4.14 (1.36)	4.61 (1.31)	4.43 (1.33)	4.00 (1.53)

Panel B: ANOVA – Effect of *Gender* on *Overlap* in *Strong Past Performance*^a

Source of Variation	df	Mean Square	<i>F</i>	<i>p</i>
<i>Gender</i>	3	0.35	0.26	0.85
Error	87	1.34		
Total	90			

Panel C: ANOVA – Effect of *Gender* on *PredictivePower* in *Strong Past Performance*^a

Source of Variation	df	Mean Square	<i>F</i>	<i>p</i>
<i>Gender</i>	3	1.79	0.93	0.43
Error	87	1.92		
Total	90			

Notes

- ^a 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* and *PredictivePower* questions; they did not receive any subordinate performance information or make judgments of potential or causal attributions.
- ^b 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?"
- ^c We measured participants' beliefs about the predictive power of past performance for future potential with this question, anchored on 1 = Strongly disagree and 7 = Strongly agree: "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?"

TABLE 6
Supplemental Analysis – Attributions and Judgments of Potential for Weak Performers

Panel A: Mean (Standard Deviation) of *Potential* in *Weak Past Performance* ^a

	<i>Gender</i> ^a		
	<i>Male</i> n = 25	<i>Female</i> n = 21	<i>Neutral</i> n = 23
<i>Ability</i> ^b	25.80 (13.59)	42.05 (19.38)	27.30 (12.55)
<i>Potential</i> ^c	3.48 (1.05)	3.71 (0.78)	3.35 (1.11)

Panel B: ANOVA – Effect of *Gender* on *Ability* in *Weak Past Performance* ^a

Source of Variation	df	Mean Square	<i>F</i>	<i>p</i>
<i>Gender</i>	2	1774.50	7.60	0.01
Error	66	233.51		
Total	68			

Panel C: ANOVA – Effect of *Gender* on *Potential* in *Weak Past Performance* ^a

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

Notes

- ^a 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*.
- ^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. In the *Weak Past Performance* conditions, higher values indicate a belief that a lack of ability is a more important driver of the observed weak performance.
- ^c We measured participants’ judgments 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?*”
-