Gender Bias through Recategorization of Financial Analysts

Robert Bloomfield  
$rjb9@cornell.edu$  
SC Johnson College of Business  
Cornell University  
Ithaca, NY 14853  
Tel: +1 (607) 255-9407

Kristina Rennekamp*  
$kmr52@cornell.edu$  
SC Johnson College of Business  
Cornell University  
Ithaca, NY 14853  
Tel: +1 (607) 255-0500

Blake Steenhoven  
$bas362@cornell.edu$  
SC Johnson College of Business  
Cornell University  
Ithaca, NY 14853  
Tel: +1 (907) 350-0878

Scott Stewart  
$sds58@cornell.edu$  
SC Johnson College of Business  
Cornell University  
Ithaca, NY 14853  
Tel: +1 (607) 255-9177

March 2019

*Corresponding Author.

We are grateful for input from Scott Asay, Natasha Bernhardt, Sanjeev Bhojraj, Rebecca Fender, Melissa Ferguson, Bob Libby, Kun Liu, and workshop participants at The University of Iowa, Tilburg University, and the 2018 ABO Conference. We acknowledge support provided by CFA Institute, CFA Society Boston, CFA Society New York, and CFA Society Chicago in distributing experiment materials. This paper reflects the views of the authors and does not represent the official views of CFA Institute or societies.
Gender Bias through Recategorization of Financial Analysts

ABSTRACT

We present 179 investment professionals with a scenario that manipulates whether a male or female analyst persists in pitching a stock pick after it has been voted down. Respondents evaluate analysts as less promotable when they do not persist, but only if the analyst is female. Results are consistent with categorization theory, which suggests evaluators shift their categorization of non-persistent women closer to the category of “woman,” and away from the category of “analyst,” while attributing the behavior of non-persistent men to contextual features. Analysis of free-response questions confirm that the unexpected behavior was a predominant focus in performance evaluations of women, while for men focus was mostly restricted to competence-related factors. Semi-structured interviews with 13 senior investment professionals provide additional support for the role of expectations and categorization heuristics on promotion decisions. Our findings shed light on factors that may contribute to the investment industry’s “leaky pipeline” for women.

JEL Codes: M40, M41, M49, M51

Keywords: performance evaluation, subjective performance measures, gender bias, analyst characteristics

Data Availability: Contact the authors.
I. INTRODUCTION

Subjective measures are often necessary in performance evaluation to account for the ambiguity in objective measures or to incorporate factors that are difficult to measure precisely (Bol 2008). As a result, innumerable factors could be included in a single evaluation, adding significant complexity to the evaluation task and allowing flexibility in the selection and weighting of performance information. We examine how evaluators cope with this complexity by relying on the mechanism of categorization. Categorization allows evaluators to simplify the process of identifying role-relevant information by establishing a set of characteristics associated with the role and selectively attending to expected behaviors (Feldman 1981). Unexpected behavior can cause evaluators to recategorize the people they are evaluating, especially if those people are notably atypical in ways that can explain the behavior, such as being female in a male-dominated profession.

To examine how recategorization affects the evaluation of people who exhibit unexpected behaviors and atypical characteristics, we examine performance evaluation for research analysts, a male-dominated role. In male-dominated settings, like asset management, expected behavior is generally characterized by “agentic” features stereotypically associated with men, like assertiveness, forcefulness, and persistence. In contrast, “communal” features like cooperativeness, deference, and sensitivity, associated with stereotypes of women, are inconsistent with expectations of analysts (Heilman 2012). When an analyst deviates from expected characteristics, evaluators who rely on categorization can shift their categorization of the analyst toward another category that might explain the behavior (Feldman 1981). We predict that the availability of “female” as a salient category that can explain a lack of persistence will increase the likelihood of recategorization when female analysts behave in unexpected ways.
Because this alternative explanatory category is unavailable for men, unexpected behavior by male analysts will not shift categorization.

We test this prediction using a 2 x 2 between-subjects experiment with 179 professional investors sourced from four conferences held by the CFA Institute or member societies.\(^1\) Participants read a scenario describing a research analyst being considered for promotion at a hypothetical hedge fund. In the scenario, the analyst presents a “high conviction” buy recommendation at a weekly investment meeting and the recommendation is narrowly voted down. We manipulate whether the analyst responds with expected behavior (by persisting with the recommendation) or unexpected behavior (by dropping the recommendation). We also manipulate whether the analyst in the scenario is male (John) or female (Jennifer). After reading the scenario, participants indicate the extent to which the facts provided help or hurt the analyst’s case for promotion.

Consistent with our predictions, we find that participants evaluate analysts who drop the recommendation more negatively than analysts who persist, but only when the analyst is female. We use independent coding of participants’ free responses describing the most important factors for the evaluation and find additional evidence that our results are driven by recategorization of the analysts after they exhibit unexpected behavior. Responses are more likely to mention the unexpected behavior (lack of persistence) and less likely to discuss the analyst’s competence when the analyst is female than when the analyst is male, consistent with greater ease of recategorizing female analysts. Reflecting attribution of the unexpected behavior to other contextual features, the mention of these items did not differ between male analysts who dropped the recommendation and male analysts who persisted with the recommendation. We provide

\(^{1}\) Approval to use human participants has been granted by the Institutional Review Board.
further evidence that this effect is driven by recategorization using a supplemental experiment with Executive MBA students. While these students have sufficient knowledge of accounting and finance to evaluate the recommendation described in the scenario, most have little to no experience in asset management and therefore have less rigid expectations of analyst behavior and less contextual knowledge. As a result, we expect that they will be less likely to identify behavior as discrepant and less capable of attributing identified discrepancies to contextual features. Consistent with recategorization requiring this institutional knowledge, we find no differences in performance evaluations across our four conditions with Executive MBA student participants.

Finally, we conduct semi-structured interviews with thirteen experienced asset managers who have substantial institutional knowledge and experience in performance evaluation. These interviews allow us to validate the assumptions underlying our theory and provide additional depth to our findings. Respondents confirm that there is substantial ambiguity in the performance evaluation of research analysts and that subjective measures are often used. In addition, interviewees largely validate that analysts are expected to be persistent, both generally and as described in our experimental materials. They also shed light on how ambiguity in performance evaluation and expectations in a male-dominated culture contribute to gender bias in the promotion of research analysts. In particular, they describe how expectations of aggressive behavior force women to choose between violating expectations of analysts and violating gender expectations, both of which can have negative consequences. Respondents also provide additional insights into the antecedents and consequences of gender bias, which expand the scope of our study and may guide future research on this important issue.
Our findings contribute to the managerial accounting literature on the subjectivity in performance evaluation (Baker, Gibbons, and Murphy 1994; Hayes and Schaefer 2000; Datar, Kulp, and Lambert 2001; Gibbs et al. 2004; Bol 2008). In particular, our study extends the literature examining how subjectivity facilitates bias in performance evaluation (Moers 2005; Bol 2011, 2008). Our findings suggest that in evaluations involving subjectivity, categorization can result in biased attributions of analyst behavior. Holding constant the qualifications of the analyst, details of the stock recommendation, and outcome of the presentation, we show that attributions of the same behavior can differ based on the availability of categories which can be used to explain the behavior. While we examine the effect of gender in a male-dominated industry, we expect that these effects are likely relevant to performance evaluation of other under-represented groups as well.

We also contribute to the broader literature on gender bias in performance evaluation. Prior research has examined obstacles to hiring and promotion in male-dominated fields due to a perceived lack of fit between the requirements of the job and the descriptive norms for women (Eagly and Karau 2002). Many of these studies show that women who exhibit agentic behaviors that are expected more of men than women are viewed as unlikeable and unkind, a phenomenon known as “backlash” (Rudman and Fairchild 2004). Our study complements this work by showing that women who fail to exhibit agentic behaviors that are expected of their profession but not of women are viewed as a poor fit for the job. That women were punished for a single communal behavior in this setting may partially explain both the scarcity of women in senior asset management positions, as well as recent survey findings that female CFA® charterholders are higher in achievement-orientation than members of the general public (Adams, Barber, and
Odean 2016), and our own interview finding that female investment professionals are more uniformly persistent than their male counterparts.

The remainder of the paper is organized as follows. In Section II, we present our theory and hypotheses. In Section III, we describe our experimental design and main results. Additional analyses, including independent coding of free responses and our supplemental experiment, are discussed in Section IV. Our semi-structured interviews are detailed in Section V. Section VI concludes.

II. THEORY AND HYPOTHESES

Subjectivity in Performance Evaluation

An expansive literature in managerial accounting documents the role of subjective factors in performance evaluation. Consistent with this literature, Farnsworth and Taylor (2006) find that compensation packages for portfolio managers are based more on subjective and discretionary factors than objective and formula-based factors. Findings from two recent surveys suggest that subjective factors also play an important role in the performance evaluation of research analysts. Brown et al. (2015) survey sell-side analysts, finding that three of the most important determinants of compensation are industry knowledge, accessibility and/or responsiveness, and professional integrity, all of which require subjective judgment. While analyst rankings and broker votes are also viewed as important, other objective measures such as the profitability of stock recommendations, success at generating commissions, and accuracy and timeliness of earnings forecasts are among the factors viewed as least important. Brown et al. (2016) conduct a similar survey with buy-side analysts. Although respondents indicate stock recommendations are the most important determinants of compensation, more subjective factors like professional integrity, accessibility and/or responsiveness, and industry knowledge are also viewed as
important. These findings make clear that performance evaluations of research analysts, both buy-side and sell-side, rely in large part on subjective factors.

Research within and outside of accounting has examined the use of subjective measures in evaluations of employee performance (Bommer et al. 1995; Gibbs et al. 2004). Much of this research focuses on the insufficiency of purely objective measures in evaluating and incentivizing performance, particularly when they are noisy, subject to manipulation, or when important characteristics are unobservable (Baker, Gibbons, and Murphy 1994; Hayes and Schaefer 2000; Datar, Kulp, and Lambert 2001; Gibbs et al. 2004; Bol 2008). The experienced asset managers in our semi-structured interviews, discussed in Section V, provide additional insight into the reasons subjective performance measures are used to evaluate research analysts. Consistent with the notion that subjective measures are used to account for noise in objective measures, one respondent noted, “It’s really hard to evaluate somebody just based on their history […] of accurate picking or not. Because they could make the right call, and yet get the wrong result. The call is based on probabilities, and the results of course are based on what happened.”

While this type of ambiguity illustrates the value of subjective measures, it also presents a potential channel for bias in the performance evaluation of research analysts. Feldman (1981) describes the appraisal process as “a complex and cyclical process” involving recognition of and attention to relevant information, organization and storage of the information, recall of the information, and integration of the information into judgments. The complexity of this process as well as cognitive limitations faced by appraisers provide ample opportunity for error and bias in performance evaluation. For example, Bol (2011) finds that in subjective performance evaluations, the cost of gathering information and the strength of the relationship with
subordinates are positively associated with centrality bias (i.e., insufficient variance in performance ratings), and leniency bias (i.e., unwarranted inflation in performance ratings). Other studies demonstrate how subjectivity allows for bias in evaluations both through the ambiguity in interpretation and relative weighting of available information. (Ittner, Larcker, and Meyer 2003; Moers 2005; Bol and Smith 2011).

**Categorization in Performance Evaluation**

The complexity of the appraisal process can also lead to an increased use of heuristic cues, tools which help simplify the information space and decision process. One such tool is categorization, whereby expectations are used to identify, organize, recall, and integrate performance information into decisions (Feldman 1981). The categorization process involves assigning individuals to various categories based on the salience of particular attributes and contextual features and using these categories as a lens to interpret information about their behavior (Srull and Wyer 1979). When faced with myriad potential cues, categorization allows appraisers to reduce the information space by relying on characteristics of the individual’s category and selectively attending to category-consistent information. As a result, this information is most easily recalled and is therefore more likely to be factored into the subsequent performance evaluation (Feldman and Lynch 1988). While this tool can be effective in simplifying the appraisal process, the role of subjective judgment throughout the categorization process, as described below, can introduce and amplify bias in performance evaluation.

Individuals are generally categorized based on similarity of their attributes to those of prototypical members of categories (Tversky 1977). As individuals are complex, they can belong to many categories simultaneously, with variation due to context as well as characteristics of the perceiver. The effect of categorization on interpretation of information as consistent or
inconsistent with expectations and the resulting effect on performance evaluation depend on the salience of the category (Feldman 1981). For example, a fund manager may simultaneously categorize a research analyst based on her occupation, gender, age, race, personality, sexual orientation, and marital status. While any or all of these categories may color how the fund manager interprets information about the analyst, the influence of each category is determined by the fund manager’s prototype of these categories and contextual features which make a particular characteristic more salient.

In a performance evaluation setting, an evaluator assesses an individual’s “fit” with a role by comparing their characteristics and behaviors to that of a typical person in that role. For research analysts, this involves assessing the similarity between the analyst and the prototypical analyst. Characteristics of the evaluator’s prototypical analyst include demographic characteristics like age, gender, and education as well as personality characteristics the evaluator has observed among research analysts. While these demographic and personality characteristics likely vary based on an evaluator’s unique experience and characteristics, a concurrent working paper by Adams, Barber, and Odean (2016) documents demographic information from member data for Chartered Financial Analysts (CFA® charterholders) as well as a survey of values. They report that roughly 82% of charterholders are male and, compared to the general population, charterholders on average place more value on achievement and less value on tradition, consistent with a stereotypically male value-orientation. These findings suggest that an evaluator’s prototypical research analyst, perhaps unsurprisingly, is likely to share characteristics with prototypical males and, as a result, evaluations of analyst behavior will be influenced by conformity with stereotypically masculine attributes (Heilman 2012). We provide further support
from semi-structured interviews of experienced asset managers for the assertion that stereotypically male characteristics are expected of research analysts in Section V.

While categorization can allow an evaluator to focus on expectation-consistent information, sufficiently inconsistent information requires that the evaluator resolve the discrepancy. This resolution typically involves attribution of the inconsistency either to the individual being evaluated or to other contextual features. The attribution of this inconsistency depends on the ease of recategorizing the individual into other categories for which the behavior is consistent (Feldman 1981). Put differently, if an individual exhibits surprising behavior, an evaluator will attempt to identify characteristics of the individual that explain the surprise. When such characteristics are identified, the behavior is attributed to the individual. If no such characteristics are identified, the behavior is attributed to other contextual features. Thus, the effect on an individual’s evaluation depends greatly on the ease of identifying explanatory characteristics. In the next section, we describe how the availability of a stereotypically feminine category can result in different attributions of behavior for male and female research analysts and the implications for gender bias in performance evaluation.

**Gender in Performance Evaluation**

As with other categories, gender categorization relies on the similarity between an individual and the prototypical member of a group. Therefore, the categorization of an individual as “male” or “female” depends largely on descriptive norms surrounding each gender. In contrast to prescriptive norms, which relate to how men and women should be, descriptive norms reflect perceptions of how men and women are and, as a result, perceptions of how they differ. While descriptive and prescriptive norms commonly overlap, they are not identical by definition and can have very different implications for stereotyping and resulting discrimination (Burgess and
Descriptive norms for men commonly describe “agentic” characteristics, including aggressiveness, decisiveness, and other traits associated with achievement. In contrast, descriptive norms for women denote a more social orientation, including cooperativeness, kindness, and other “communal” characteristics. While none of these characteristics carry a negative connotation in isolation, expectations of men and women can result in biased evaluations due to a perceived “lack of fit” between the gendered characteristics and the requirements of a job (Heilman 2001). In male-dominated environments, where perceptions of career success are associated with agentic traits, evaluations of women may be negatively affected due to expectations that women do not possess these traits (Eagly and Karau 2002).

In performance evaluations, particularly when information is ambiguous and subjective judgment is required, the perceived lack of fit can introduce gender bias throughout the appraisal process. As women are expected to have communal characteristics, evaluators’ attention may be directed toward behaviors that reflect these, rather than agentic characteristics. As a result, communal characteristics will be recalled more easily during subsequent performance evaluations, creating greater distance between women and the prototypically successful analyst, and reinforcing the perceived lack of fit (Heilman 2012).

The effect of this self-reinforcing bias largely depends on the categorization process. Evaluators assign individuals to numerous categories, with occupational categories made especially salient during performance evaluation. The heightened salience of features associated with occupational categories, particularly in experimental studies with limited contextual information, has been posited as an explanation for failure to identify effects of other variables in performance evaluation. However, certain contextual features may heighten the salience of other
features like gender, particularly when behavior differs from expectations of the occupation (Feldman 1981). These discrepancies and subsequent recategorizations occur when an individual’s role-incongruent features are made salient, such as the exhibition of communal behavior in a role where agentic behavior is expected (Heilman 2012).

For research analysts, there are many activities in which agentic characteristics may be expected to manifest. In addition to the technical activities directed toward research, analysts must also be effective in communicating their ideas and opinions. As one asset manager we interviewed noted, analysts “need a very outgoing, argumentative sort of approach to their process. As opposed to just being someone who just does analytical work all day, and writes reports, and then sends them on.” In these interpersonal settings, characteristics like persistence, forcefulness, and confidence are on display and provide an opportunity for analysts to either confirm that they possess these characteristics or to exhibit behavior discrepant from expectations of research analysts. If an analyst violates expectations by exhibiting communal characteristics or failing to sufficiently exhibit agentic characteristics, a recategorization process may occur so that an evaluator can resolve the perceived inconsistency between their view of the analyst as competent and the prototypically agentic view of a successful analyst (Feldman 1981).

As such, our first hypothesis is as follows:

**H1**: When research analysts exhibit communal, rather than agentic characteristics, performance evaluations will be negatively impacted.

Whether the recategorization is applied to the analyst (shifting weight to categories other than those associated with successful analysts) or by attributing the lack of agentic characteristics to contextual or environmental features depends on the relative ease of these processes. For female research analysts, the availability of gender as a category with communal descriptive norms can account for a lack of agentic characteristics. We therefore argue that the ease of
recategorizing the individual is greater for female analysts than for male analysts. This tendency is exacerbated by the salience of gender in finance resulting from the scarcity of women in the field (Adams, Barber, and Odean 2016).

Our second hypothesis is as follows:

\textbf{H2}: The negative effect of exhibiting communal, rather than agentic, characteristics will be greater for female analysts than for male analysts.

III. EXPERIMENT

Participants

Participants are 179 professional investors from four conferences held by the CFA Institute or member societies. Our theory relies on experience in finance and, as such, these participants are well-matched to our experiment. We discuss the importance of expertise in our additional analysis below. We do not collect demographic information, though we expect the demographics to be similar to those of professional investors in general.\(^2\) Because our instrument was paper-based, requesting demographics carried a risk that participants would be concerned about being identified and might revise previous responses, which would reduce the benefits of our between-subjects design. However, because our theory relies on descriptive norms, we expect demographics to have little effect on behavior. Descriptive norms are broadly consistent across demographics, including age and gender (Spence and Buckner 2000; Eagly and Karau 2002; Heilman 2012). Further, despite changes in social roles and attitudes, descriptive norms have proven resistant to change over time (Haines, Deaux, and Lofaro 2016; Croft, Schmader, and Block 2015). This is consistent with the results of similar studies analyzing stereotype-based gender biases which show no effects of evaluator gender (Rudman and Glick 1999; Lyness and

\(^{2}\) See Adams, Barber, and Odean (2016) for a recent survey of CFA® charterholders. Demographic information can also be found for buy-side and sell-side analysts in Brown et al. (2016) and Brown et al. (2015), respectively.
Heilman 2006; Madera, Hebl, and Martin 2009). Further, in our pretesting through Amazon Mechanical Turk, we did not identify effects of demographic variables. Finally, assuming our sample is demographically similar to the general population of professional investors, any demographic effects in our study will be similar to those in practice.

**Design**

Our experiment uses a 2 x 2 between-subjects design in which participants read a scenario involving a research analyst and evaluate the extent to which the scenario helps or hurts the analyst’s case for promotion to a sector specialist position. In the scenario we manipulate the analyst’s gender (male, female) and response to their recommendation being voted down (persistent, not persistent). We hold constant the characteristics of the analyst’s hedge fund, experience, and education. We manipulate the analyst’s gender by using a traditionally male or female name (John/Jennifer) and manipulating pronouns accordingly (he/she, his/her).

In the scenario, the analyst presents a stock recommendation to analysts and the portfolio manager at the hedge fund’s weekly investment meeting. We hold constant the details of the analyst’s stock recommendation and that the recommendation is narrowly voted down after a contentious discussion. We manipulate the analyst’s response as *Not Persistent* by stating that after the vote, the analyst “decided not to pursue the matter further, and did not bring it up in subsequent meetings.” In the *Persistent* condition, we state that after the vote, the analyst arranged to meet with the portfolio manager to make their case again. To reproduce the ambiguity inherent in performance evaluation for research analysts, participants are not told

---

3 Other studies show effects of participant gender on evaluations (Eagly, Makhijani, and Klonsky 1992; Eagly, Karau, and Makhijani 1995; Bowen, Swim, and Jacobs 2000). However, as noted by Bowen et al. 2000, the effects identified in these studies reflect “the tendency of men to hold more traditional role beliefs,” consistent with prescriptive norms. If our results were driven by the proportion of male respondents in our sample, this would result in a main effect of analyst gender, rather than the interaction we predict.
whether the recommended stock is added to the portfolio or any information about the performance of the stock.4

After reading the scenario, participants in the Male (Female) condition answer the following question: “In your view, do the events summarized above help or hurt John (Jennifer)’s case for promotion?” This evaluation is our dependent variable of interest and responses are provided on a scale ranging from -10 (“Strongly Hurt”) to 10 (“Strongly Help”). Participants are then asked to provide a free response indicating which factors were most important to their assessment. Finally, participants respond to questions about how future recommendations should be handled, the analyst's strengths and weaknesses, and the importance of these factors.

Procedure

Conference attendees were randomly presented envelopes with experimental materials for one of the four conditions. Written instructions inform participants that they will be asked to evaluate the performance of a financial analyst who is currently being considered for promotion at a hypothetical hedge fund. All participants receive an identical description of the hedge fund, which employs 30 analysts, who work under seven sector specialists, all under the leadership of a single portfolio manager. To convey that each recommendation is important, participants are informed that the fund’s strategy is to concentrate on a small number of positions, so each recommendation is closely monitored and factors heavily into analysts’ performance evaluation.

4 Given the high value of our professional participants, we conducted two pilot tests of our instrument using Amazon Mechanical Turk. The instrument used in these tests was similar to the one given to our professional participants, included in Appendix A, but also manipulated the analyst’s primary industry (automotive or mining), whether the stock was added to the portfolio or not, and whether the stock ultimately had a positive or negative return. As expected, we found no effect of industry and dropped this manipulation for our professional participants. Based on results of the two outcome manipulations and through conversations with experts, we determined that including this objective information failed to reproduce the ambiguity inherent in performance evaluations through which recategorization occurs in practice, as expressed by Feldman (1981). As such, these manipulations were also dropped for our professional participants.
Participants then receive a description of an analyst being considered for promotion to Industrials Sector Specialist. In all conditions, the analyst has a Master of Engineering and an MBA, has four years of work experience in a consulting group for industrials, and has covered the automobile industry at the fund for the past five years. In the Male condition, the analyst is named John and male pronouns (he, his) are used. In the Female condition, the analyst is named Jennifer and female pronouns (she, her) are used.

All participants then read details of the analyst’s recent stock pick, a “high conviction” buy recommendation for Automatic, Inc. The recommendation is based on an attractive entry point due to recent revenue declines and an upcoming launch of a new product line which the analyst determined would be highly successful and capture significant market share, leading to substantial positive stock returns. In all conditions, the analyst pitches this recommendation at the fund’s weekly investment meeting and, after a contentious discussion, the group votes narrowly against the recommendation. In the Not Persistent condition, the analyst does not pursue the recommendation further and does not present it again. In the Persistent condition, the analyst arranges to meet with the fund’s portfolio manager to make their case again. After reading this scenario, participants provide their evaluation of the analyst.

Results

Our hypotheses predict that analysts who exhibit communal characteristics by not persisting with their recommendation will be evaluated more negatively than analysts who exhibit agentic characteristics by persisting with their recommendation (H1) but that the negative evaluation will be larger when the analyst is female than when the analyst is male (H2). Panel A of Table 1 presents descriptive statistics for evaluation ratings and Panel B provides the ANOVA results. Results are presented graphically in Figure 1.
Consistent with H1, the main effect of persistence is significant (F(1,175) = 7.62, p = 0.003, one-tailed), indicating that analysts who exhibit communal characteristics by not persisting are evaluated more negatively than analysts who exhibit agentic characteristics by persisting. Further, we find support for H2, with a significant interaction of gender and persistence (F(1, 175) = 3.85, p = 0.026, one-tailed), suggesting that the negative effect of communal characteristics on evaluation is greater for female analysts than for male analysts.

We further examine the shape of the interaction through simple main effects tests presented in Table 1, Panel C. These tests confirm that the effect of persistence is significant for female analysts (t-stat = 3.41, p < 0.001, one-tailed), with female analysts evaluated worse when they are not persistent. We further find that female analysts who are not persistent are evaluated significantly more negatively than male analysts who are not persistent (t-stat = 2.25, p = 0.013, one-tailed). Consistent with evaluators relying on expectations of analysts, rather than gender categories, when congruent behavior is displayed, we do not find a difference in how persistent male and female analysts are evaluated (t-stat = 0.58, p = 0.565, two-tailed). Interestingly, we do not find a significant effect of persistence for male analysts (t-stat = 0.55, p = 0.582, two-tailed). This is consistent with greater difficulty of recategorizing male analysts leading to attribution of this behavior to contextual or environmental features. We provide further support for this explanation in the next section.
IV. ADDITIONAL ANALYSES

Analysis of Free Responses

To provide further evidence for the reasons underlying the differential evaluations, we examine participants’ free responses. According to our theory, behavior that is discrepant from the expectations of research analysts will lead to a recategorization process, either of the individual or of contextual features. When no unexpected behavior is exhibited, we expect evaluations to reflect agentic factors, consistent with the characteristics expected of research analysts. When discrepant behavior like a lack of persistence is exhibited, the salience of gender as an explanatory factor for the behavior facilitates recategorization of female analysts relative to male analysts, for which this category is not available. As such, we expect evaluations of female analysts who are not persistent to differ from those of persistent analysts and male analysts who are not persistent. Specifically, we expect that these evaluations will focus more on persistence, the characteristic related to the individual’s discrepant behavior.

Two independent coders were recruited to code the factors included in participants’ free responses. Coders were blind to hypotheses and, to ensure that coders were blind to conditions, mentions of the analyst’s name were changed to “Analyst X” and all pronouns were replaced nongendered pronouns (e.g. they, their). Free responses were coded based on whether they mentioned the quality of the stock pick or the analyst’s research (“Analyst’s Competence”) or the analyst’s level of conviction, persistence, or assertiveness from their actions taken after the recommendation (“Analyst’s Persistence”). As we are specifically interested in whether recategorization causes evaluative criteria to differ, coders are asked to indicate if “factors are

---

5 Coders were also asked to identify whether responses included the analyst’s internal beliefs about the quality of their recommendation and the analyst’s ability to persuade others. These items received the lowest counts of the four items coded and as such, we do not discuss them further.
referenced in responses, regardless of whether the reference is positive, negative, or neutral.” Agreement between coders for the initial coding is 83% and we find a Kappa coefficient of 0.651, indicating “substantial” agreement. Coders met independently to resolve coding of the remaining items (Landis and Koch 1977).

Table 2 presents the coding results for “Analyst’s Competence” and the proportion of free responses including the quality of the stock pick or the analyst’s research are presented graphically in Figure 2. Results of the nominal logistic regression in Panel B generally align with the pattern of results for Experiment 1, with a significant difference in mentions of competence in responses by Persistence (Chi-square = 3.04, p = 0.041, one-tailed), consistent with analysts who exhibit behavior discrepant from expectations being recategorized and evaluated based on different criteria. While we again find no main effect of Gender (Chi-square = 0.78, p = 0.377, two-tailed), we document a significant interaction of Gender and Persistence (Chi-square = 7.79, p = 0.003, one-tailed), suggesting that the effect of this recategorization on evaluative criteria is stronger for female analysts than for male analysts.

Individual tests in Panel C also align with simple main effects tests in Experiment 1. Competence was included in the evaluation of persistent female analysts significantly more frequently than for non-persistent female analysts (Chi-square = 10.36, p = 0.001, one-tailed). We also find that competence is mentioned less frequently when the non-persistent analyst is female than when the non-persistent analyst is male (Chi-square = 6.67, p = 0.005, one-tailed).
There was no difference in the frequency with which competence was discussed between persistent and non-persistence male analysts (Chi-square = 0.58, p = 0.446, two-tailed) or between persistent male and female analysts (Chi-square = 1.95, p = 0.163, two-tailed). These results collectively suggest that analysts who behave consistent with expectations are evaluated based on their competence and analysts who violate these expectations are evaluated based on other criteria, particularly when another salient category, like gender, is available.

Table 3 presents the coding results for “Analyst’s Persistence” and the proportion of free responses including the analyst’s persistence, specifically related to actions taken after the investment meeting, are presented graphically in Figure 3. Results of the nominal logistic regression in Panel B demonstrate a pattern of results roughly the inverse of those in Experiment 1. Consistent with expectations, we find a significant difference in mentions of the analyst’s persistence, conviction, or assertiveness by Persistence (Chi-square = 5.07, p = 0.012, one-tailed). We do not find a significant main effect of Gender (Chi-square = 2.44, p = 0.118, two-tailed), but find a marginally significant interaction of Gender and Persistence (Chi-square = 2.50, p = 0.057, one-tailed), suggesting that the focus on the lack of persistence is greater for female analysts than for male analysts.

-----------------------------------------
Insert Figure 3 about here
-----------------------------------------

-----------------------------------------
Insert Table 3 about here
-----------------------------------------

Individual tests in Panel C provide further evidence on the use of this evaluative criteria. While persistence was included in the evaluation of non-persistent female analysts significantly more frequently than for persistent female analysts (Chi-square = 8.29, p = 0.002, one-tailed), we
do not find a difference between persistent and non-persistent male analysts (Chi-square = 0.21, p = 0.647, two-tailed) or between persistent male and female analysts (Chi-square = 0.00, p = 1.000, two-tailed). Further, persistence is mentioned significantly more frequently when the non-persistent analyst is female than when the non-persistent analyst is male (Chi-square = 5.82, p = 0.008, one-tailed). These results suggest that behavior inconsistent with expectations of research analysts, specifically a lack of persistence, is incorporated into evaluations of female analysts, but not male analysts.

**Supplemental Experiment**

To provide further evidence that the effects we document are driven by categorization, we conduct a supplemental experiment to examine whether these results hold for individuals without experience as research analysts. Recategorization in this setting requires strong expectations of research analysts such that discrepant behavior can be identified as well as knowledge of the setting to allow for attribution of discrepant behavior either to the evaluated analyst or to contextual features. As individuals without experience as research analysts are unlikely to have this knowledge, we would not expect discrepant behavior and subsequent recategorization to affect their evaluations.

We conduct our supplemental experiment with 98 Executive MBA students as participants. The experiment is identical both in content and delivery to Experiment 1. We use MBA students as participants because they have sufficient knowledge of accounting and finance to understand the materials, but do not have experience as research analysts and thus are unlikely to have the experience and institutional knowledge required for identification of discrepant behavior and subsequent recategorization to occur.

Consistent with Experiment 1, we do not find a significant main effect of Gender.
In contrast with Experiment 1, we do not find a significant main effect of Persistence (F(1,94) = 0.77, p = 0.381, two-tailed, untabulated) or a significant interaction between Gender and Persistence (F(1,94) = 0.01, p = 0.915, two-tailed, untabulated). Follow up simple effects tests also reveal no differences between conditions (all p > 0.45, untabulated). These results provide evidence that experience as a research analyst is important to our findings and are consistent with the role of expectations and contextual knowledge leading to identification of discrepant behavior and recategorization.

V. SEMI-STRUCTURED INTERVIEWS

To provide additional evidence on the role of subjectivity in the evaluation of research analysts as well as the expectations held by evaluators, we conduct semi-structured interviews with thirteen experienced asset managers. We interview six female and seven male asset managers with an average of 24.5 years of investment experience. Eleven participants have held the role of research analyst and eleven have held the role of portfolio manager, confirming that these participants have extensive institutional knowledge as well as experience in hiring and performance evaluation. Participants also have diverse experience within the investment industry, with six participants specializing in equity and seven specializing in fixed-income. Participants also have experience in firms of different sizes, measured by assets under management (AUM), with eleven having worked in large firms (over $100B AUM), four having worked in mid-sized firms (between $1B and $100B AUM), and three having worked in small firms (under $1B AUM). In sum, these experienced asset managers have substantial and diverse industry experience.

Given our interest in participants’ unique experiences, we chose to conduct semi-structured interviews wherein specific questions were prepared, but interviewers were free to
deviate from the prepared questions to pursue interesting ideas in participants’ responses. As in Hirst and Koonce (1996), we did not want to limit the discussion to our prepared questions, as our goal was to obtain a better understanding of performance evaluation. The semi-structured approach allowed us to obtain rich data by encouraging participants to further develop and contextualize their responses. Each interview was approximately 30 minutes in length and was conducted using video conferencing software, with two authors conducting the interview remotely and one author facilitating and recording the interview in person.6

The list of prepared questions for the interviews is provided in Appendix B. The prepared questions were developed to elicit expectations of research analysts (Q1), to confirm the role of subjective factors in performance evaluation (Q2), to confirm the expected course of action for the analyst in our experimental instrument (Q3), and to receive general views from evaluators on the obstacles to promotion for female analysts (Q4). When participants did not specifically refer to persistence unprompted, we asked about the importance of this characteristic as well (Q2.1) to ensure that views on this characteristic were provided. We summarize some of the themes in responses and provide illustrative quotes below.7

**Characteristics of Analysts**

To validate the assumptions underlying our theory, we asked respondents to describe the expected characteristics of research analysts and how these expectations factor into performance evaluations. Participants generally confirmed the view that expectations of analysts include agentic characteristics like persistence, forcefulness, and confidence.

---

6 Due to availability constraints, three of the thirteen interviews were conducted in person by a single author. While these differences likely produced minor differences, including different follow-up questions from interviewers, the purpose of these interviews was to provide additional qualitative evidence and we do not attempt to quantify findings from these interviews. As such, we do not exclude these interviews from our discussion.

7 Each quote is followed by an indicator for the respondent, including the respondent’s gender, an identifier number based on order, and the number of the question to which they were responding. As an example, a quote in response to Question 3 from the second female participant we interviewed would be followed by [F23].
I would think analysts working in investment management, where there is a decision point, may be more aggressive than if you were analyzing sequencing DNA on a longer run project. Every day there’s a, “fish or cut bait,” on every idea. [F11]

[T]hey also have to have an outgoing personality, because they need to sell their ideas, or sell their work to the various fund managers. [...] there’s point-counterpoint challenges to various assessments, and so forth. So they need a very outgoing, argumentative sort of approach to their process. As opposed to just being someone who just does analytical work all day, and writes reports, and then sends them on. [M11]

I think the most important attribute is to have confidence. So if I’ve done the first few things well—like really tested my hypothesis, and really done my homework, and really sure about what I’m saying—you can really speak with confidence. [F31]

Respondents also provided reasons why the association between agentic characteristics and research analysts originated and how these expectations are perpetuated.

[P]eople tend to have a vision of how—whoever had that job before—that’s how they envision that job being done, how that person did it. And they think that that makes sense. So, “They were successful, they did it this way, and they looked that way, and they carried themselves this way, and behaved a certain way.” […] I don’t really think it’s always front of the mind, but in the back of the mind you’re like, “Oh, well that’s what that person—that person should continue to be like that, who does that role.” [F54]

But yeah, I mean, a culture gets perpetuated, and a culture that rewards gregariousness, and aggressiveness, and sort of willingness to stick your neck out there, and not be shy, and err on the side of over-communication—because that’s the better side to err on—you know, a culture that rewards that, it gets perpetuated. [F22]

According to participants, deviations from these expected characteristics can result in concerns about an individual’s ability to succeed in this role, reflecting the “lack-of-fit” described by Heilman (2001).

I think you need to like to interact with people. I think it’s challenging to be too—you can be a little bit of an introvert, but I think it’s challenging to be too much of an introvert and to be successful. […] I know some analysts that still don’t like talking in the big setting, right? They get nervous in front of a lot of people. And like, that’s ok. But you can’t do that, and then also not be able to interact and engage in the smaller settings as well. [F51]

I interviewed someone the other day for a job here, and it was, sort of a—it wasn’t an analyst job—but it was a job where they were going to advise clients on their investment
portfolios, and work with clients. And my five minute snap judgement impression that I
gave myself after meeting this person was, “Way too timid and shy.” [F21]

Consistent with our experimental findings, several respondents noted that the consequences of
deviating from expectations can be more severe for female analysts than for male analysts,
suggesting this as one cause of the industry’s “leaky pipeline” for women.

*I think you see that across financial firms, where you have a whittling down over time of
the women who are in the organization. So if all the senior leaders have similar
characteristics, whatever they may be, then people who are most successfully coming up
through the ranks of the organization are most likely to mirror those characteristics.*
[F14]

*If a shy woman comes in, you just think, “Well she’s just a shy—she’s a woman.” If it’s a
male, maybe you think, “Maybe he’s so smart, he can’t communicate or something.”*
[M64]

Participants further described how these different standards affect promotion outcomes, as
reflected in the characteristics of male and female portfolio managers. Whereas male portfolio
managers are viewed as more characteristically heterogeneous, women promoted to portfolio
manager were described as uniformly assertive and aggressive.

*I think—at least from my years at [investment firm]—the male portfolio managers, you
do [...] get a much more broader cross-section, at least along the spectrum of shyer
personality vs. more aggressive personality. But all the women I knew in portfolio
managers roles: very aggressive. Definitely. Yeah. Outspoken from an average person’s
standpoint, maybe too much so.* [F21]

*In some ways you think of them as more male than the males—sort of the women that
make it to portfolio manager. [...] you don’t have to be ruthless to be a portfolio
manager; you can be just like a normal person, and a nice person. But in general, the
people who make it to that level are pretty tough people.* [M24]

These responses are consistent with the achievement orientation of female analysts documented
in Adams, Barber, and Odean (2016) and with our experimental findings. Whereas persistence
did not affect evaluations of male analysts in our experiment, investment professionals viewed
non-persistent women as less promotable than persistent women. As such, we would expect less
aggressive women to be underrepresented relative to less aggressive men in these roles. Further, we note that the perception that female portfolio managers as “more male than the males” may exacerbate the consequences of non-agentic behavior for female analysts. As the category of “female portfolio manager” may be more characteristically agentic than that of “male portfolio manager,” a violation of expectations may be more salient for female analysts than for male analysts.

Response to Experimental Scenario

After validating our theoretical assumptions, we provided participants with a description of the scenario described in our experimental instrument and asked how a research analyst would be expected to respond when their recommendation is voted down. Respondents generally indicated that the expected response would be to follow up on the recommendation, confirming that dropping the recommendation would represent behavior incongruent with expectations.

I think you would most likely regroup, repackage, and re-pitch it. Or pitch it in a more customized way. [F13]

I like to be persistent. So yeah, I’ve had people with dissenting opinions, and if I have really strong conviction, I’m going to keep presenting it. And I think that’s how over time portfolio managers get to trust you. Because if I don’t believe in my conviction, or in my investment conclusion, then really why should they? [F63]

It could be that they just felt that they didn’t have enough information, and so perhaps the analyst should rethink, go back, and take another look at their idea or recommendation, and perhaps present it again if it’s one that they still have a lot of conviction over […] I remember a couple times doing that, just going back and rethinking, looking for more information, trying to present a different way of considering a specific stock, then at least over time ultimately being successful—assuming you still believe in the story. [M13]

Participants noted that failing to persist with a recommendation has negative connotations, even describing experiences where analysts who exhibited these characteristics were fired.

Some cases we’ve had instances where the analyst just never really followed up. And they didn’t follow up, and we’d encourage them to follow up, and they didn’t follow up, we’d encourage them to follow up, and they didn’t. And then it’s time to let them go. [M53]
If you never mention it again, it probably means it didn’t go up, so you don’t want to draw attention to the fact that you recommended this name, and it didn’t go up, right? [F53]

One participant specializing in fixed-income noted the opposite expectation, but still indicated that a dropped recommendation is a negative signal, especially if the behavior reflects a pattern.

Well, I guess it’s a two part answer. First, I would expect them to drop it. Unless they have other information that they need to share with you, but if they’ve already shared all their information, then they should drop it. Number two is a more fundamental question: If this happens a lot, why do you have that analyst? What’s the point? [M73]

Some respondents noted that a dropped recommendation was not necessarily a negative reflection on the analyst. Contextual features like portfolio objectives, sector preferences, and fund characteristics could be used to justify a dropped recommendation.

So it depends. I mean, if you’re an analyst covering 100 companies, in your scenario, and you present this one idea, and it’s narrowly voted down [...] you just push it aside, even if you feel really strongly about it, and you go on to your other 99 companies, or whatever. I mean, if you’re only covering 10 companies, and you feel really strongly about this, and you have a lot of these character traits that we’re talking about, being aggressive, you come back at it. [...] And if you do that, and you do that successfully, and you’re right, that’s when you really get rewarded. [F23]

And then obviously there’s also the committee thinking, “Gee, this is an interesting idea, but it’s not in a sector—based on where we see the economy or market going—that we’re really interested in right now. So, you know, great idea. Thanks. However, this is banking, and we prefer manufacturing,” [M13]

In sum, respondents confirmed that the generally expected behavior of analysts in our experimental scenario would be to follow up on the recommendation and that the behavior of the non-persistent analysts was incongruent with expectations. Respondents also indicate how dropping a recommendation can be attributed to the analyst’s “lack of fit” or to contextual features. This is consistent with the findings of our free response analysis, in which performance evaluations of non-persistent women focused on persistence, whereas those of non-persistent men focused on the quality of the stock recommendation.
The Role of Communication Style in Performance Evaluation

As our theory asserts that categorization is a response to subjectivity in evaluations, we asked participants about the use of qualitative factors in performance evaluations of research analysts. Respondents described communication style as one of the most important qualitative factors in performance evaluation.

"You can be a great stock-picker and have really bad communication skills. But, if you’re working for an investment firm, communication skills are definitely really important, both verbal and written." [F21]

"If you sit in your office and send out an email saying, “Hey I like this stock,” but never talk to anyone, and never—right? It’s harder to be rewarded for that. No one expects you to get it right all the time. But being in front of people and being vocal is important." [F53]

Many participants noted that agentic characteristics like persistence, assertiveness, and confidence are on display in analysts’ communication, with several respondents describing successful communicators as those who “pound the table.”

"There is a level of pounding of the table that just has to happen. So I can walk downstairs to the trading room and I can say, “I think X company, the return on the bonds are going to come back to par. They’re trading at 70. They’re a survivor.” And they’re not going to do it unless I tell them three or four times, and I list out why, and I have to kind of pounding the table when it’s a risky investment." [F61]

"There’s a fair degree of self-promotion that has to take place to get yourself noticed. Obviously the results have to back that up, because there’s a lot of self-promotional people that just don’t seem to be able to pick stocks all that well. [...] the person who was outspoken, somewhat self-promotional, clearly had an advantage. Not all the time, but clearly had an advantage." [M12]

Although some respondents indicated that these agentic characteristics in communication are necessary for success, others expressed concern that these characteristics are over-weighted in performance evaluation or may be used to mask poor underlying performance.

"I mean, we definitely have analysts who will get in there and pound the table on ideas, and I would suspect that it’s difficult not to get caught up in the enthusiasm that an..."
analyst portrays when they’re making their point. [...] I don’t think there’s any way not to have some form of bias in your receptivity to that. [M52]

But in person, I would say conviction is given a greater weight than clarity of presentation. And so the one thing I haven’t really said is anything about accuracy. And in convincing portfolio managers, accuracy doesn’t seem to matter all that much. [M22]

Some participants noted how communication presents a tightrope for female analysts who view assertiveness as necessary for success, but may face backlash when these characteristics manifest in their communication. Others expressed frustration with a perceived inability to account for differences in communication styles in performance evaluation.

I think some of it goes back to [...] that aggressive personality. And it’s like, women get in trouble because to be good in this field, you do have to have aggressive personality. But to be too aggressive, you’re kind of like a bitchy woman, and who wants to be friends with you? [F24]

I don’t know of any process that can account for how people may perceive those different characteristics. They have an impact, I know that they have an impact—different communication styles, different degrees of assertiveness, different forms of communication—but I don’t know how to measure them. I don’t know how we measure them. [F12]

In total, responses reflect that analysts are expected to exhibit agentic characteristics, especially in communication of research ideas, due to expectations resulting from and perpetuated by a male-dominated industry. As such, failing to show persistence, as in our experimental scenario, represents discrepant behavior, which generally results in a negative evaluation. However, there are contextual factors to which the behavior can be attributed, though female analysts may not receive the benefit of these attributions, particularly when the discrepant behavior reinforces a “lack of fit.”

VI. CONCLUSION

We examine one channel through which ambiguity in performance evaluation can allow for bias in promotion decisions for research analysts. In our main experiment with senior
investment professionals, we find that failing to exhibit persistence, an expected characteristic of analysts, results in lower performance evaluations for female, but not for male, analysts.

Analysis of participants’ free responses, a supplemental experiment with EMBA students, and semi-structured interviews with experienced asset managers provide convergent evidence for our hypothesis that the effect is driven by violation of expectations and a resulting shift from the category of “analyst” toward the category of “woman,” but no comparable shift toward another category for men. Our results suggest that ambiguity in the evaluation process allows unexpected behavior by male analysts to be attributed to contextual features, while the same behavior by female analysts is viewed as indicative of a “lack of fit.”

Our findings complement prior studies documenting negative evaluations of women who exhibit agentic characteristics and barriers to promotion for women in male-dominated industries (Heilman et al. 2004; Lyness and Heilman 2006) by examining consequences of communal behavior in a male-dominated industry. In this setting, we show that female analysts who behave communally receive worse evaluations, while male analysts are not evaluated negatively for the same behavior. We reconcile these results with literature describing “backlash” against individuals who deviate from gender-stereotypical behavior (Rudman and Fairchild 2004) by providing evidence that the availability of gender as a category for female analysts leads to perceptions that the unexpected behavior is evidence of a “lack of fit.” As no such category exists for male analysts, the discrepant behavior is attributed to contextual features or simply ignored.

Our results appear consistent with admonitions to “lean in,” particularly in male-dominated roles and industries (Sandberg 2013). This interpretation is supported by the achievement-orientation of female CFA® charterholders (Adams, Barber, and Odean 2016), and
with impressions of the experienced asset managers we interviewed, who indicate that women who are promoted are “more male than the males.” However, given the impact of a single lack of persistence, our findings suggest that the reality may be more akin to “lean in always, or else.” The effects we report could contribute to the relative scarcity of women in senior levels, the agentic features associated with successful women in finance, and findings that promotions for female managers are associated with higher performance evaluations, suggesting women are held to stricter standards than men (Lyness and Heilman 2006).

The professional benefits of persistence for women are likely specific to several aspects of our setting and elicitation. First, we study judgment in an industry that is particularly agentic. A concurrent working paper shows that in group recruiting events for public accounting firms, female applicants are punished for acting in a stereotypically “male” manner (Fanning, Williams, and Williamson, 2018). We suspect that the collaborative nature of auditing results in the category of “auditor” reflecting more communal characteristics than that of “analyst,” especially in social settings where gender differences are expected to manifest and comparisons are more salient. Second, we study judgments of individuals already employed as analysts. We suspect that women face more difficulty than men in establishing themselves into the category of “analyst” in the first place. This difficulty is reflected in recent findings by Botelho and Abraham (2017) that in online investment platforms with uncertain information, gender bias is observed in the initial selection of recommendations.

Finally, we note our primary dependent variable asked respondents about the “promotability” of analysts. Research shows that women who succeed in male sex-typed positions can be perceived as deficient in desirable qualities of women and are viewed less

---

8 This quote was provided by one of the male asset managers in our semi-structured interviews in response to Question 4 from our script (see Appendix B).
favorably as a result (Heilman and Okimoto 2007). Thus, while participants viewed a failure to persist as making women less promotable in this particular scenario, the unabated persistence required for a ‘lean in always’ approach could lead women to be viewed as unlikable, and make colleagues less likely to offer professional opportunities other than promotion, such as information sharing, invitations to professional and social events that offer ‘face time’ with clients and superiors, and assignment to managerial and supervisory duties that are demanded at the highest levels of the investment profession.
The purpose of this study is to better understand how financial analysts are evaluated. To do so, we will ask you to evaluate the performance of a financial analyst being considered for promotion at a hypothetical hedge fund.

The Hedge Fund
BTF runs a $10 billion equity hedge fund. BTF typically holds only 25-30 stocks at a time, so every stock position matters. The firm employs 30 analysts, and portfolios are managed by 7 sector specialists under the leadership of a single Portfolio Manager, Bill Thomas. Thomas is considering promoting some high-performing analysts to sector specialists.

Due to the fund’s strategy of concentrating on a small number of positions, each recommendation is monitored closely and weighs heavily on the evaluation of analyst performance. The firm therefore tracks the outcome of recommendations presented at weekly meetings held to evaluate changes in the funds’ positions.

The Analyst
Jennifer covers the automobile industry and is being considered for promotion to Industrials Sector Specialist. She earned a Master of Engineering from Michigan and an MBA from Columbia. Following 4 years in a consulting group covering industrial companies, she has been a research analyst for 5 years at the firm.

Recent Stock Pick
Jennifer presented a “high conviction” buy recommendation on Automatic, Inc., a small-cap stock in the auto components industry.

She noted that Automatic, Inc. had been facing revenue declines over the last few years due to increasing competition but was preparing to launch a new signature product line. This line offered unique new features based on patented proprietary technology, though its development required a significant capital outlay.

After studying patents and other technical literature on the technology developed by Automatic and their competitors, along with market research on features of the products, Jennifer determined that Automatic Inc. would have a highly successful launch, capture significant market share, and see a significant stock price increase.

Presentation
Jennifer presented this stock idea at the weekly investment meeting to analysts and the portfolio manager. After a contentious discussion, the group voted narrowly against her recommendation.

Jennifer decided not to pursue the matter further, and did not bring it up in subsequent meetings.

Appendix A. Experimental Materials

The purpose of this study is to better understand how financial analysts are evaluated. To do so, we will ask you to evaluate the performance of a financial analyst being considered for promotion at a hypothetical hedge fund.

The Hedge Fund
BTF runs a $10 billion equity hedge fund. BTF typically holds only 25-30 stocks at a time, so every stock position matters. The firm employs 30 analysts, and portfolios are managed by 7 sector specialists under the leadership of a single Portfolio Manager, Bill Thomas. Thomas is considering promoting some high-performing analysts to sector specialists.

Due to the fund’s strategy of concentrating on a small number of positions, each recommendation is monitored closely and weighs heavily on the evaluation of analyst performance. The firm therefore tracks the outcome of recommendations presented at weekly meetings held to evaluate changes in the funds’ positions.

The Analyst
Jennifer covers the automobile industry and is being considered for promotion to Industrials Sector Specialist. She earned a Master of Engineering from Michigan and an MBA from Columbia. Following 4 years in a consulting group covering industrial companies, she has been a research analyst for 5 years at the firm.

Recent Stock Pick
Jennifer presented a “high conviction” buy recommendation on Automatic, Inc., a small-cap stock in the auto components industry.

She noted that Automatic, Inc. had been facing revenue declines over the last few years due to increasing competition but was preparing to launch a new signature product line. This line offered unique new features based on patented proprietary technology, though its development required a significant capital outlay.

After studying patents and other technical literature on the technology developed by Automatic and their competitors, along with market research on features of the products, Jennifer determined that Automatic Inc. would have a highly successful launch, capture significant market share, and see a significant stock price increase.

Presentation
Jennifer presented this stock idea at the weekly investment meeting to analysts and the portfolio manager. After a contentious discussion, the group voted narrowly against her recommendation.

Jennifer decided not to pursue the matter further, and did not bring it up in subsequent meetings.

Appendix A presents the experimental materials for the Female and Not Persistent condition. Other conditions were identical except that in Male conditions, the analyst’s name is John and male pronouns are used, and in the Persistent condition, the last sentence reads “After the vote, Jennifer (John) arranged to meet with Bill Thomas, the Portfolio Manager, to make her (his) case again.”
## Appendix B. Prepared Questions for Semi-Structured Interviews

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>What would you say are the characteristics of a typical analyst? How do these characteristics manifest in analysts’ day to day activities?</td>
</tr>
<tr>
<td>Q2</td>
<td>Do you think analysts are evaluated solely based on their stock picks, or are other factors also considered? What other factors are important?</td>
</tr>
<tr>
<td>Q2.1</td>
<td>How important are characteristics like assertiveness, aggressiveness, and persistence?</td>
</tr>
<tr>
<td>Q3</td>
<td>Suppose an analyst presents a recommendation at a firm meeting and the recommendation is narrowly voted down. What are the next steps they would take? What would you usually do in this situation?</td>
</tr>
<tr>
<td>Q4</td>
<td>The CFA Institute has documented that a lot of women are not promoted from analyst to fund manager. Why do you think that might be?</td>
</tr>
</tbody>
</table>
VII. References


Figure 1 plots responses from professional investor participants in Experiment 1 to the question “In your view, do the events summarized above help or hurt John’s (Jennifer’s) case for promotion?” Responses are provided on a scale ranging from -10 (“Strongly Hurt”) to 10 (“Strongly Help”). Our experiment manipulates the behavior of the analyst (Persistent or Not Persistent) and the analyst’s gender (Male or Female) in a 2×2 between-participants design. A total of 179 participants provide an evaluation for the analyst.
Figure 2 plots the proportion of free responses from participants in Experiment 1 including mention of the analyst’s competence by condition. Two independent raters were instructed to code responses =1 “if the explanation references the quality of the stock pick or the analyst’s research, analysis, due diligence, knowledge, expertise, or qualifications” and =0 otherwise. Raters met to resolve any differences in coding. Our experiment manipulates the behavior of the analyst (Persistent or Not Persistent) and the analyst’s gender (Male or Female) in a 2×2 between-participants design. A total of 168 participants provide a free response with their evaluation of the analyst.
Figure 3 plots the proportion of free responses from participants in Experiment 1 including mention of the analyst’s persistence by condition. Two independent raters were instructed to code responses =1 “if the explanation references the analyst’s level of conviction, persistence, or assertiveness, specifically as it relates to the actions taken after the meeting” and =0 otherwise. Raters met to resolve any differences in coding. Our experiment manipulates the behavior of the analyst (Persistent or Not Persistent) and the analyst’s gender (Male or Female) in a 2×2 between-participants design. A total of 168 participants provide a free response with their evaluation of the analyst.
### TABLE 1
Experiment 1
Descriptive Statistics and Analysis of Variance

**Panel A: Impact of Scenario on Analyst Promotability, Mean (Standard Deviation) [n]**

<table>
<thead>
<tr>
<th>Persistence</th>
<th>Male</th>
<th>Female</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persistent</td>
<td>0.410</td>
<td>0.978</td>
<td>0.714</td>
</tr>
<tr>
<td></td>
<td>(4.717)</td>
<td>(4.595)</td>
<td>(4.632)</td>
</tr>
<tr>
<td></td>
<td>[39]</td>
<td>[45]</td>
<td>[84]</td>
</tr>
<tr>
<td>Not Persistent</td>
<td>-0.128</td>
<td>-2.208</td>
<td>-1.179</td>
</tr>
<tr>
<td></td>
<td>(4.609)</td>
<td>(4.094)</td>
<td>(4.458)</td>
</tr>
<tr>
<td></td>
<td>[47]</td>
<td>[48]</td>
<td>[95]</td>
</tr>
<tr>
<td>Overall</td>
<td>0.116</td>
<td>-0.667</td>
<td>-0.291</td>
</tr>
<tr>
<td></td>
<td>(4.639)</td>
<td>(4.607)</td>
<td>(4.626)</td>
</tr>
<tr>
<td></td>
<td>[86]</td>
<td>[93]</td>
<td>[179]</td>
</tr>
</tbody>
</table>

**Panel B: Analyst Evaluation ANOVA**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>25.45</td>
<td>1</td>
<td>25.45</td>
<td>1.26</td>
<td>0.264</td>
</tr>
<tr>
<td>H1: Persistence</td>
<td>154.14</td>
<td>1</td>
<td>154.14</td>
<td>7.62</td>
<td>0.003†</td>
</tr>
<tr>
<td>H2: Gender x Persistence</td>
<td>77.94</td>
<td>1</td>
<td>77.94</td>
<td>3.85</td>
<td>0.026†</td>
</tr>
<tr>
<td>Error</td>
<td>3539.56</td>
<td>175</td>
<td>20.23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Panel C: Analyst Evaluation Simple Effects**

<table>
<thead>
<tr>
<th>Effect on Evaluation</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of Persistence for Female Analysts</td>
<td>3.41</td>
<td>0.000†</td>
</tr>
<tr>
<td>Effect of Persistence for Male Analysts</td>
<td>0.55</td>
<td>0.582</td>
</tr>
<tr>
<td>Effect of Gender for Persistent Analysts</td>
<td>0.58</td>
<td>0.565</td>
</tr>
<tr>
<td>Effect of Gender for Non-Persistent Analysts</td>
<td>2.25</td>
<td>0.013†</td>
</tr>
</tbody>
</table>

† One-tailed (or equivalent), given our directional predictions.

Table 1, Panel A presents descriptive statistics for professional investor participants’ responses by condition in Experiment 1 to the question “In your view, do the events summarized above help or hurt John’s (Jennifer’s) case for promotion?” Responses are provided on a scale ranging from -10 (“Strongly Hurt”) to 10 (“Strongly Help”). Panel B presents hypothesis test results for the effect of Persistence and Gender on evaluations and Panel C presents simple main effects. Our experiment manipulates the behavior of the analyst (Persistent or Not Persistent) and the analyst’s gender (Male or Female) in a 2×2 between participants design. A total of 179 participants provide an evaluation for the analyst. All p-values provided are two-tailed, with the exception of the tests of our hypotheses. These directional expectations are indicated in the table with a dagger, as noted above.
### TABLE 2
Independent Coding of Free Responses, Experiment 1 – Analyst’s Competence

#### Panel A: Responses Coded as Referring to Analyst’s Competence

<table>
<thead>
<tr>
<th>Persistence</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent</td>
<td></td>
<td>14 out of 38</td>
<td>22 out of 42</td>
<td>36 out of 80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>37%</td>
<td>52%</td>
<td>45%</td>
</tr>
<tr>
<td>Not Persistent</td>
<td></td>
<td>19 out of 42</td>
<td>9 out of 46</td>
<td>28 out of 88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45%</td>
<td>20%</td>
<td>2%</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>33 out of 80</td>
<td>31 out of 88</td>
<td>64 out of 168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41%</td>
<td>35%</td>
<td>8%</td>
</tr>
</tbody>
</table>

#### Panel B: Nominal Logistic Regression

<table>
<thead>
<tr>
<th></th>
<th>Chi-Square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.78</td>
<td>0.377</td>
</tr>
<tr>
<td>Persistence</td>
<td>3.04</td>
<td>0.041†</td>
</tr>
<tr>
<td>Gender x Persistence</td>
<td>7.79</td>
<td>0.003†</td>
</tr>
</tbody>
</table>

#### Panel C: Chi-Square Tests of Association

<table>
<thead>
<tr>
<th></th>
<th>Chi-Square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent vs Non-Persistent Female Analysts</td>
<td>10.36</td>
<td>0.001†</td>
</tr>
<tr>
<td>Persistent vs Non-Persistent Male Analysts</td>
<td>0.58</td>
<td>0.446</td>
</tr>
<tr>
<td>Male vs Female Persistent Analysts</td>
<td>1.95</td>
<td>0.163</td>
</tr>
<tr>
<td>Male vs Female Non-Persistent Analysts</td>
<td>6.67</td>
<td>0.005†</td>
</tr>
</tbody>
</table>

†One-tailed (or equivalent), given our directional predictions.

Table 2, Panel A presents the percentage of free responses from participants in Experiment 1 including mention of the analyst’s competence by condition. Two independent raters were instructed to code responses =1 “if the explanation references the quality of the stock pick or the analyst’s research, analysis, due diligence, knowledge, expertise, or qualifications” and =0 otherwise. Raters met to resolve any differences in coding. Panel B presents results of a nominal logistic regression for the effects of Persistence and Gender on mentions of the analyst’s competence and Panel C presents chi-square tests of association for differences between conditions. Our experiment manipulates the behavior of the analyst (Persistent or Not Persistent) and the analyst’s gender (Male or Female) in a 2×2 between-participants design. A total of 168 participants provide a free response with their evaluation of the analyst. All p-values provided are two-tailed, with the exception of the tests of our hypotheses. These directional expectations are indicated in the table with a dagger, as noted above.
### TABLE 3

**Independent Coding of Free Responses, Experiment 1 – Analyst’s Persistence**

#### Panel A: Responses Coded as Referring to Analyst’s Persistence

<table>
<thead>
<tr>
<th>Persistence</th>
<th>Male</th>
<th>Female</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent</td>
<td>10 out of 38</td>
<td>11 out of 42</td>
<td>21 out of 80</td>
</tr>
<tr>
<td></td>
<td>26%</td>
<td>26%</td>
<td>6%</td>
</tr>
<tr>
<td>Not Persistent</td>
<td>13 out of 42</td>
<td>26 out of 46</td>
<td>39 out of 88</td>
</tr>
<tr>
<td></td>
<td>31%</td>
<td>57%</td>
<td>4%</td>
</tr>
<tr>
<td>Overall</td>
<td>23 out of 80</td>
<td>37 out of 88</td>
<td>60 out of 168</td>
</tr>
<tr>
<td></td>
<td>29%</td>
<td>42%</td>
<td>36%</td>
</tr>
</tbody>
</table>

#### Panel B: Nominal Logistic Regression

<table>
<thead>
<tr>
<th></th>
<th>Chi-Square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>2.44</td>
<td>0.118</td>
</tr>
<tr>
<td>Persistence</td>
<td>5.07</td>
<td>0.012†</td>
</tr>
<tr>
<td>Gender x Persistence</td>
<td>2.50</td>
<td>0.057†</td>
</tr>
</tbody>
</table>

#### Panel C: Chi-Square Tests of Association

<table>
<thead>
<tr>
<th></th>
<th>Chi-Square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent vs Non-Persistent Female Analyst</td>
<td>8.29</td>
<td>0.002†</td>
</tr>
<tr>
<td>Persistent vs Non-Persistent Male Analyst</td>
<td>0.21</td>
<td>0.647</td>
</tr>
<tr>
<td>Male vs Female Persistent Analyst</td>
<td>0.00</td>
<td>1.000</td>
</tr>
<tr>
<td>Male vs Female Non-Persistent Analyst</td>
<td>5.82</td>
<td>0.008†</td>
</tr>
</tbody>
</table>

† One-tailed (or equivalent), given our directional predictions.

Table 3, Panel A presents the percentage of free responses from participants in Experiment 1 including mention of the analyst’s persistence by condition. Two independent raters were instructed to code responses =1 “if the explanation references the analyst’s level of conviction, persistence, or assertiveness, specifically as it relates to the actions taken after the meeting” and =0 otherwise. Raters met to resolve any differences in coding. Panel B presents results of a nominal logistic regression for the effect of Persistence and Gender on mentions of the analyst’s persistence and Panel C presents chi-square tests of association for differences between conditions. Our experiment manipulates the behavior of the analyst (Persistent or Not Persistent) and the analyst’s gender (Male or Female) in a 2×2 between participants design. A total of 168 participants provide a free response with their evaluation of the analyst. All p-values provided are two-tailed, with the exception of the tests of our hypotheses. These directional expectations are indicated in the table with a dagger, as noted above.