DO TRENDS MATTER? THE EFFECTS OF DYNAMIC PERFORMANCE TRENDS AND PERSONALITY TRAITS ON PERFORMANCE APPRAISALS

JOSEPH A. SCHMIDT

Edwards School of Business, University of Saskatchewan

Acknowledgements

I would like to thank Christopher Poile, Chelsea Willness, and Dionne Pohler for their helpful comments on prior versions of this manuscript. This research was funded in part by the the Social Sciences and Humanities Research Council of Canada (# 430-2014-00383).

DO TRENDS MATTER? THE EFFECTS OF DYNAMIC PERFORMANCE TRENDS AND PERSONALITY TRAITS ON PERFORMANCE APPRAISALS

ABSTRACT

Two studies were conducted to understand how people make overall performance judgments based on dynamic performance trend information and the personality characteristics of ratees. University athletes were sampled in Study 1 and the results showed that improving performance trends resulted in higher appraisals of task performance. Contrary to previous experimental research, raters did not use trend information to make attributions about the targets' effort or other behavioral characteristics. There were also interactions between performance trends and personality: performance trends were positively associated with task performance ratings for players with high extraversion and low agreeableness, while trends were unrelated to ratings for players at the opposite end of the continuum for these traits. The second study was an experiment designed to test the potential theoretical mechanisms that explained the effects observed in Study 1. The results showed that raters used performance trend information to derive task performance ratings, while they used personality information to derive ratings of citizenship behavior. Attributions about employee effort and ability were based on both performance trends and personality. The results also indicated that raters engaged in more deliberative (controlled) cognitive processing when the target's personality and performance trend were incongruent, which may explain the interaction effects observed in Study 1. Implications for theories of social cognition and performance appraisal are discussed.

Keywords: performance evaluation; personality; longitudinal trends; growth modeling

DO TRENDS MATTER? THE EFFECTS OF DYNAMIC PERFORMANCE TRENDS AND PERSONALITY TRAITS ON PERFORMANCE APPRAISALS

Performance variability remains a simultaneously frustrating and intriguing puzzle. Sports fans lament about the inconsistent performance of their favorite athletes, managers wonder why last year's star employee is this year's dud, and analysts struggle to explain why formerly successful CEOs frequently fail at new companies. Researchers have established that individual performance varies over time, and that there are individual differences in patterns of performance variability (e.g., Hofmann, Jacobs, & Baratta, 1993; Hofmann, Jacobs, & Gerras, 1992; Rabbitt, Osman, Moore, & Stollery, 2001). Although this research helps us understand within-person performance variance, we know less about how raters incorporate variability into their overall performance judgments (Fisher, 2008). Reb and Greguras (2008) argue that deficiencies in our understanding of how raters integrate dynamic behavioral information into performance judgments may provide an important explanation for the relatively low correlation between objective performance metrics and performance ratings (see the meta-analysis by Bommer, Johnson, Rich, Podsakoff, & MacKenzie, 1995).

In spite of the ubiquity of performance appraisals, both employees and managers are often dissatisfied with the process (Murphy & Cleveland, 1995). Employees perceive that ratings are biased and unfair; front-line managers feel uncomfortable in the dual role of coach and evaluator; and senior managers often question if the performance review process actually improves future employee and firm performance (DeNisi & Sonesh, 2011). Researchers have also noted the inherent biases in the performance appraisal process, with some going so far as to suggest that research using performance appraisals as the primary criterion is untrustworthy (e.g., Osterman, 2007). It is no wonder that companies such as General Electric and Adobe are

abandoning the traditional performance review process in order to provide employees with more frequent and forward-looking feedback (Nisen, 2015; Rock & Jones, 2015).

Given the problems with performance reviews, it is tempting to stop research and practice on the topic. However, valid performance judgements are required to make many important personnel decisions processes – not the least of which are promotion, compensation, and career development (DeNisi & Pritchard, 2006). While acknowledging the problems plaguing performance reviews, Murphy and Cleveland (1995) argued that it is reckless to abandon the practice entirely. Others have argued that employee reactions to appraisals should continue to be an important focus of research (DeNisi & Murphy, 2017) given that employees who are subject to transparent and inclusive performance management systems are more likely to be satisfied with the appraisal results and have lower turnover intentions (Taylor, Tracy, Renard, Harrison, & Carroll, 1995). Assuming that raters can accurately encode and utilize dynamic performance trend information, explicit consideration of behavioral trends in the performance judgment process may increase accuracy of the ratings and enhance employees' perceptions of fairness and satisfaction with the process.

The limited research that has been conducted about the effects of dynamic performance trends on performance ratings has produced valuable insights. Barnes, Reb, and Ang (2012) showed that performance trends predicted subsequent compensation in the National Basketball Association, suggesting that general managers used trend information in their player performance evaluation processes. Other scholars found that raters use dynamic trends as heuristics to allocate employees into performance categories or prototypes (DeNisi & Stevens, 1981; Lee & Dalal, 2011; Reb & Cropanzano, 2007) and to make attributions about the behaviors of employees (Reb & Greguras, 2010). For example, employees with improving

performance trends are perceived to have higher ability and work harder than those with declining trends (Reb & Greguras, 2010; Rudolph, Harari, & Nieminen, 2015).

Although these studies have made insightful contributions to our knowledge about how raters incorporate dynamic information into overall judgments, extant theory on performance trends may contradict more established theories of social cognition. Performance trend theories propose that observers use trend information as a heuristic to form impressions about the target's characteristics (Reb & Greguras, 2010; Rudolph et al., 2015). This type of information processing schema suggests that, in field settings, raters must encode a series of performance episodes as they unfold over time, interpret the trend pattern from the series of observations, and then form impressions about the target's personality and unobserved behaviors (e.g., ability, effort) based on perceptions about the performance trend. On the other hand, Srull and Wyer's (1989) model of social information processing suggests that people tend to quickly classify others into a general category or prototype (e.g., outgoing), often based on initial impressions or observing only a small sample of behavior. The prototype is applied to interpret subsequent behavior and "fill in the gaps" for unobserved behavior. These two theoretical perspectives appear to be at odds and seem to provide competing propositions about how raters derive performance judgments based on the ratees' personal characteristics and performance trend information.

In reality, performance evaluators have opportunities to observe ratees' behavioral and objective performance information over a relatively long period of time. Current theory does not explain how raters attend to and integrate these different performance cues into overall evaluations. Will evaluators process behavioral information via performance trend heuristics, person prototypes, or some combination of the schemas?

Another critique of performance trend information processing theory is that, with the exception of Barnes et al. (2012), the theory was developed based largely on experimental studies (DeNisi & Stevens, 1981; Lee & Dalal, 2011; Reb & Cropanzano, 2007; Reb & Greguras, 2010; Rudolph et al., 2015; Scott & Hamner, 1975). The experiments were well-designed and experimental controls are required to establish causality; however, consistent with prior critiques about experimental performance evaluation research (e.g., Spence & Keeping, 2011), the study participants lacked information about interpersonal interactions, ratee personality, or the social context in which ratings occurred. This may explain the potential inconsistencies between theories about performance trend information processing and other theories of social cognition.

Given these concerns, the current research was designed to answer the following questions with two studies: (1) Are raters capable of attending to and processing dynamic performance trend information in field settings? and (2) Do performance trends influence attributions that observers make about targets' personal characteristics and behaviors, or does the arrow point the other way? Study 1 was conducted with a field sample of university football players where I tested how raters (supervising coaches in this sample) integrated game performance trend information and player personality to derive overall ratings of task performance and other behaviors (e.g., "coachability" and work ethic). Study 2 was an experiment designed to test the causal effects of ratee personality and performance trends on performance ratings as well as the theoretical mechanisms that explained the interaction effects observed in Study 1. That is, did ratee likableness or inconsistency between ratee personality and performance trends cause raters to engage in more deliberative information processing?

Theoretical Background

Given the considerable cognitive demands required to process social information, judges tend to act as "cognitive misers" (Harris, Ipsas, & Schmidt, 2008) by applying heuristics and schemas to efficiently integrate behavioral observations. Srull and Wyer (1989) propose a social cognitive model that clarifies the steps involved in the impression formation process. Simply stated, judges integrate initial observations about a target's behavior to form a "general evaluative concept," which is a pre-existing category that guides the attention, encoding, and recall of future behavior. If behavior is inconsistent with the general concept, judges may engage in deeper information processing to better understand the behavior and explore the validity of the evaluative concept. Unless engaged in deeper or more controlled information processing, people retrieve the general evaluative concept, rather than specific behaviors, to make judgments about others. Automatic information processing based on general evaluative concepts sometimes leads to biased judgments, especially when judges apply the concept to make inferences about the target's unobserved traits or behaviors (i.e., halo effects or illusory correlations: Nisbett & Wilson, 1977; DeNisi, Cafferty, & Meglino,1984). Similar theories of social information processing have been called person impression or prototype schemas in the performance appraisal literature (e.g., Favero & Ilgen, 1989; Foti & Lord, 1987).

Raters in field settings are likely to apply prototype schemas (Murphy & Cleveland, 1995) and the extent to which they do so reduces the likelihood that they will encode performance trends. Trends unfold over long periods of time and require substantial effort to collect and interpret. In contrast, general evaluative impressions are based on a quick and efficient process of integrating and categorizing observed behaviors. Since performance trends are not fully observed or understood until the end of the evaluation period, the trends are likely to

be interpreted in light of the general person impression formed by the judge earlier in the evaluation period. Some research supports these arguments showing that judges use less than 25% of the available information to make attributions about others (Major, 1980) and that raters may be less likely to search for additional information after they form an initial impression of the target (see DeNisi et al., 1984). This line of reasoning suggests that observations of trait-relevant behaviors will have stronger effects on general person impressions than performance trends, and that trends may be inferred or interpreted in light of other traits.

On the other hand, prior research has shown that raters use trend information as a heuristic to derive overall evaluations (Reb & Cropanzano, 2007) and further apply trends to make attributions about the ability and effort of ratees (DeNisi & Stevens, 1981; Reb & Greguras, 2010; Rudolph et al., 2015). This may suggest that, rather than using a prototype schema, raters are processing information via a script, which is defined as a pre-existing cognitive category that outlines the structure and temporal sequence required to accomplish a goal (Foti & Lord, 1987). In other words, raters are capable of encoding the sequence of performance episodes and then classify the sequence as typical of a certain level of performance, trait, or behavior.

Although researchers have induced participants to process performance information via scripts (Foti & Lord, 1987), there are reasons to question if raters naturally apply script schemas in the field. As previously mentioned, lab studies have been used to develop theory about how raters process performance trend information. The participants were presented with a summary of the performance trend information, and were asked to make attributions about the targets' other traits. Unable to observe the behavior they were asked to evaluate, participants could only use performance trend information to make trait judgments. In real world settings, judges are

able to observe trait-relevant behaviors directly and the dynamic trends are observed as they occur. When both types of information are available, there is less need to rely on inferential leaps to make trait judgments. Given the potential conflicts between theories about performance trend information processing and other theories of social cognition, one of the main objectives of this research is to understand if raters in field settings detect and apply performance trend information to make attributions about ratees' task performance and other behavioral criteria (e.g., ability and effort).

Interactions between Performance Trends and Personality Traits

If judges are primarily focused on performance trend information, it is unlikely that interactions between trends and ratee personality will account for any variance in overall performance judgments – trends will be an important cause of task and behavioral ratings and ratee personality information will be ignored. However, if judges are applying a person-prototype schema, then ratee personality may be the primary cause of ratings and objective performance trends will have weak effects. Given that judges in field settings will have many opportunities to observe both the ratees' behavior and objective performance throughout the evaluation period, either scenario seems unlikely and extant theory is silent about how judges process performance trend information that may conflict with the target's other behaviors.

Theoretical imprecision about how raters integrate personality and performance trend information can be understood with an example based on trait conscientiousness. People who demonstrate high levels of conscientiousness tend to be diligent, organized, hard-working, and prudent, so it is not surprising that this trait has been linked to competent performance in many types of roles (e.g., Barrick, Mount, & Judge, 2001; Penney, David, & Witt, 2011). Given that an upward performance trend is consistent with the characteristics of high effort and motivation

(Reb & Greguras, 2010), it should also be consistent with high conscientiousness. Models of social cognition and the cognitive models of performance appraisal argue that characteristics that are consistent allow the rater to engage in *automatic processing* (Ilgen & Feldman, 1983; Srull & Wyer, 1989). When using automatic processing, the rater follows a well-learned stimulus-response link. Here, improving employees with conscientious personalities would get higher appraisals than declining employees with non-conscientious personalities. However, if the personality trait is inconsistent with the other observed data, this unusual mismatch may trigger a deeper *controlled processing* of the information. Improving but non-conscientious employees, or declining but conscientious employees, would cause raters to pause and think carefully about their assessments.

When engaged in controlled processing it is unclear how the rater will categorize the ratee and evaluate the competing information. Suppose the rater had a non-conscientious ratee with an improving performance trend. Perhaps the rater will discount the non-conscientious behavior and evaluate based on the performance trend. Or, perhaps the rater will allow the non-conscientious behavior to bias their assessment and discount the actual performance. As such, the second objective of this research was to test interactions between personality characteristics and performance trends to understand how judges integrate and weigh conflicting behavioral and performance trend information in their performance assessments.

I focused on testing interactions with the traits of extraversion, agreeableness, and conscientiousness as these traits may have important effects on the general impressions formed by judges and influence how judges process dynamic trend information. People with high extraversion and low agreeableness may engage in more status-striving and confrontational behaviors in a group context (Hogan & Holland, 2003). Research has shown that some judges

may be particularly attuned to these traits in others in order to predict the tone of future social interactions (Erez, Schilpzand, Leavitt, Woolum, & Judge, 2015). Conscientiousness has been shown to be a consistent predictor of task performance and it is possible that diligent, prudent, and goal-oriented behaviors of ratees may be a salient aspect of raters' general person impressions in competitive athletic and workplace environments.

STUDY 1 METHOD

Context

Study 1 was conducted with a field sample of university football players. Although there may be differences between athletic and employee samples, there is a rich tradition of researchers studying organizational phenomena within the athletic context (e.g., Barnes & Morgeson, 2007; Barnes et al., 2012; Day, Gordon, & Fink, 2012; Lyons, Hoffman, & Michel, 2009), due in part to the extensive performance and behavioral data that is available from such samples.

Another advantage of the sample is that the performance evaluations were conducted only for research purposes and were not used to inform administrative decisions or provide the players with developmental feedback. Murphy and Cleveland (1995) define this type of evaluation as a performance *judgment* and differentiate it from a performance *rating*, which is shared with the people being evaluated and other members of the organization. Evaluators may be motivated to alter performance ratings in order to reduce conflict with low performers, appear to be a competent leader, or reward the person being evaluated (Harris, 1994; Murphy & Cleveland, 1995). For these and other reasons, ratings are particularly susceptible to conscious distortion and performance judgments are likely the most appropriate criteria for initial field research on this topic.

Sample

The sample consisted of university football players from six different teams located in Western Canada. The position groups included in the study were defensive backs, defensive linemen, linebackers, quarterbacks, running backs, and receivers. Offensive linemen were not included in the analysis because game statistics were not recorded for these players.

Originally, 432 players completed a personality questionnaire, which represented a response rate of 89% based on the players listed on the roster of each team. To ensure that there were enough performance episodes to appropriately model within-person performance trends, only players who participated in at least three games were included in the analysis. Given that each player did not participate in every game or complete the physical ability testing (one of the controls), the sample size was reduced considerably when the personality questionnaires were matched to game performance, the control variables, and coach performance appraisals. The final sample consisted of 1,229 performance episodes (games) for 199 individuals. The mean age of the sample was 21.27 (SD = 2.22) and all of the participants were male.

Measures

Control Variables. Physical ability was entered as a control variable because it likely had important effects on player performance in this context. The physical ability variable was a composite calculated by averaging scores that were standardized within each position group for height, weight, vertical jump, reverse scored 40 yard run time, and a reverse scored agility run time. Some teams did not report vertical jump or agility run times for all players and those items were excluded from the composite for those players. Dummy variables for team and position were also entered as controls when testing the hypotheses to account for the time invariant effects of unobserved variables, such as team climate and coach rating tendencies.

Personality. The abbreviated 60-item form of the HEXACO-PI (Ashton & Lee, 2009) was used to assess six broad domains of player personality. Three of the domains (conscientiousness, extraversion, and openness to experience) are very similar to their counterparts in other Five Factor personality models. Two HEXACO factors contain slight rotational differences compared to Five Factor models. Specifically, the HEXACO agreeableness domain includes (lack of) anger, which is normally an aspect of the emotional stability domain in Five Factor models. Sentimentality is captured in the emotionality domain of HEXACO, while it composes the agreeableness domain of Five Factor models. The sixth domain, honesty-humility, is unique to this model and offers a number of theoretical and empirical advantages over other measures (Ashton & Lee, 2009). People who score on the upper end of this trait tend to, "avoid manipulating others for personal gain, feel little temptation to break rules, are uninterested in lavish wealth and luxuries, and feel no special entitlement to elevated social status" (Lee & Ashton, n.d.). The Cronbach Alphas were .70, .66, .79, .77, .74, and .75 for honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness, respectively.

Game Performance. Objective performance data was taken from individual game statistics for each of the eight regular season games. A number of statistics relevant to each position were standardized within each position group and then averaged to create composite performance score in a similar manner to previous research using football players (Lyons et al., 2009: see Table 1 for the game statistics included for each position). This performance measure can then be directly compared across all positions. For example, a linebacker and a quarterback with a standardized score (z-score) of one indicates that both players were performing at a similar level (approximately the 84^{th} percentile) compared to their peers in the same position.

Performance Evaluation. At the end of the season, the position coaches rated each player's overall performance on a five-item scale developed by Piedmont, Hill, and Blanco (1999). The ratings were completed on a five-point scale (1 = unsatisfactory to 5 = exceptional) and the five performance dimensions included: "Coachability (the player's ability to listen, learn, and apply coaches' instructions)," "Athletic ability (the amount of athleticism the player exhibited)," "Game Performance (how well the player performed overall in games)," "Team player (the ability of the player to get along and mesh with teammates on and off the field)," and "Work ethic (amount of effort and commitment the player dedicated to the team, himself and the coaches)."

Procedure

At the beginning of the season, the author traveled to each of the team sites and administered the personality and demographic questionnaire. At the end of the season, longitudinal team and player performance data were collected from the Canadian Interuniversity Sport web site. The coaches were mailed performance questionnaires at the end of the season and returned them in self-addressed envelopes.

Analysis

Latent Growth Curve Modeling. Player performance trends over time were examined with latent growth curve modeling conducted with Mplus 7.3 (see Byrne, 2012 and Ployhart & Hakel, 1998 for reviews). This procedure is a variant of structural equation modeling where performance at each time period load on to latent performance trend intercept and slope constructs. The loadings were centered such that the intercept represented performance in the first game of the season. To test the research questions, the constructs were grand-mean centered

and performance assessments were regressed on the appropriate intercept, slope, personality, and latent interaction terms.

STUDY 1 RESULTS

Table 2 displays the descriptive statistics and correlations among the study variables. Ratings of game performance were consistently associated with actual game performance at each time period. Some personality traits were also associated with the coach ratings. Honesty-humility and conscientiousness were positively correlated with the coachability dimension, and conscientiousness was also positively associated with ratings of work ethic. Notably, extraversion was negatively correlated with the coachability and team player dimensions, suggesting that coaches may have formed negative impressions of extraverts possibly because these players were more likely to socialize and less attentive to the coaches' instructions. These initial results indicate that coaches account for both objective performance and personality information to make performance judgments.

I modeled the linear, quadratic, and cubic latent growth curves to assess how different trend lines fit the data. Neither the mean intercept in the linear model ($M_{\text{INTERCEPT}}$ = .135, p = .084) nor the mean slope were statistically significant (M_{LINEAR} = .011, p = .506), indicating that there were no consistent patterns in performance trends between players. However, both the intercept and slope variance were statistically significant ($Var_{\text{INTERCEPT}}$ = .486, p = .001 and Var_{LINEAR} = .014, p = .048), which suggests that there was between-player variance in performance at the beginning of the season (indicated by the significant intercept variance) and performance trajectory over the course of the season (indicated by the slope variance). For the quadratic trend model, neither the slope mean ($M_{\text{QUADRATIC}}$ = -.006, p = .362) nor the slope

variance ($Var_{QUADRATIC}$ = .000, p = .716) were statistically significant. The cubic trend model did not converge. Therefore, I used the linear latent growth model to test the research questions.

To investigate the research questions, I regressed the five performance dimensions rated by the coaches on the performance trend and player personality traits. As displayed in Table 3, conscientiousness was positively associated with coachability and work ethic, extraversion was negatively associated with coachability and team player, and honesty-humility was positively associated with coachability. Finally, physical ability scores along with the game performance intercept and slope were positively associated with coaches' ratings of athletic ability, whereas only the game performance intercept and slope were positively associated with the coaches' ratings of game performance. These results indicate that the coaches did not attribute behavioral tendencies, such as coachability and work ethic, to players based on performance trends, which is inconsistent with earlier lab study findings. However, consistent with earlier findings, coaches attributed higher (athletic) ability to players with improving performance trends.

Table 4 shows the results of the interactions between the traits of extraversion, agreeableness, and conscientiousness with the latent intercept and slope terms on game performance ratings. There were significant interactions between the latent slope term with extraversion (b = 3.87, p = .012) and agreeableness (b = -3.42, p = .022), while the interaction with conscientiousness was not statistically significant. The regions of significance for the interaction effects are displayed in Figures 1 and 2, which show that the performance trend-rating slopes were statistically significant at moderate to high levels of extraversion and moderate to low levels of agreeableness. The follow-up tests of simple slopes show a stronger positive association between the slope term and game performance ratings when extraversion was one standard deviation (SD) above the mean (b = 6.39, p < .001) compared to one SD below

the mean (b = 2.06, p = .222). Alternatively, the association was stronger when agreeableness was one SD below the mean (b = 6.23, p = .001) compared to one SD above the mean (b = 2.20, p = .245). These findings suggest that coaches used the longitudinal trends to make game performance ratings for players who were more likely to have had a negative impact on the social context.

STUDY 1 DISCUSSION

This study was among the first to examine how raters use longitudinal performance and personality information to derive performance appraisals in a field setting. Ascending performance trends (indicating improvement over the course of the season) were positively associated with game performance ratings. Physical ability, game performance, and personality predictors were appropriately associated with their respective performance rating dimensions. Together, the results suggest that coaches did not make attributions about players' behaviors based on performance trends. However, for the game performance rating dimension, there were interactions that suggested coaches were likely to be lenient for players who possessed "prosocial" personality traits. For those with more status-striving and confrontational personality tendencies, appraisals were closer to actual game performance information. These results are inconsistent with current theory for two reasons: (1) judges did not appear to use performance trend information to make attributions about effort and other behaviors; and (2) it appeared that performance trend information was detected and/or utilized only for targets with certain personality traits.

There are two probable explanations for the latter finding. The first is that coaches may have simply liked the players with prosocial traits and given them the "benefit of the doubt." These players received a positive evaluation even if their performance declined, suggesting that

coaches were engaging in automatic information processing for players they liked personally. The coaches may have been more inclined to give negative evaluations to dislikable players; however, supervisors prefer not to give out poor evaluations (Spence & Keeping, 2011). They also find it difficult to move from the role of motivator to evaluator, and are reluctant to evaluate the worth of other human beings (Spence & Keeping, 2011). As DeNisi et al. (1984) state, "the consequences a rater perceives a poor rating will have for a ratee may simply motivate the rater to search for reasons to give a better rating" (374). As a result, it may be that supervisors spend more time and effort processing the information of the unlikeable subordinate. This may explain why the performance trend had a stronger effect on game performance ratings for those displaying non-prosocial personality traits. In other words, the coaches may have worked harder to justify their appraisals for players who detracted from the social context.

The second explanation is that raters tend to spend more time and provide more accurate appraisals for ratees who are atypical for a cognitive evaluation category, compared to ratees who are prototypical of a category (Favero & Ilgen, 1989). When an unlikeable subordinate displays a positive trend, the supervisor may have a more difficult appraisal. The unlikeable subordinate's recent performance improvement is not typical of the category, causing the supervisor to spend more time processing the information to come to a decision (Favero & Ilgen, 1989). Other research suggests that category inconsistent information is simply recalled more easily and has a stronger effect on evaluations (see Ilgen, Barnes-Farrell, & McKellin, 1993). Although plausible, both of these explanations remain untested. Therefore, a second experimental study was designed to replicate the Study 1 main effects in a lab context and determine whether likability or category inconsistency influenced the depth of information processing and performance rating accuracy.

STUDY 2

Overview

This study extended previous experimental performance trend studies in that target personality (likability) and the depth of information processing (cognitive load) were manipulated in addition to performance trends. The study was a split-plot design where participants were randomly assigned to two between-subjects conditions (high versus low cognitive load) where they evaluated the performance of four fictional employees (i.e., four within-subjects conditions: likable personality with improving performance trend, dislikable personality with improving trend, likable personality with declining trend, dislikable personality with declining trend).[1]

STUDY 2 METHOD

Sample

To understand how people with employment experience process and evaluate performance information, 158 participants were recruited from the Canadian online survey panel, Probit. Most of the participants were employed full-time (83.5%), 7.0% were part-time employees, and 9.5% were self-employed. A total of 55.7% of the sample had either completed a bachelor's degree or a vocational/technical diploma, 13.3% had completed a graduate degree (Masters or PhD), 27.2% had completed high school, and the remaining 1.3% had some high school education. Males were 59.5% of the sample.

Procedure

Participants read an introductory paragraph asking them to assume the role of a sales manager who was required to supervise and manage the performance of sales representatives for

[1] The Appendix contains information on the pilot testing procedures for this study.

an information technology services company. They were then provided with a brief sales representative job description that was taken from the O*Net database (www.onetonline.org). Finally, participants were informed that four new sales representatives were recently hired and more experienced employees provided feedback about the new employees' behaviors.

Each within-subjects condition consisted of a vignette that contained the experienced employees' descriptions of the new employee's behaviors (i.e., personality traits) and the employee's performance trend over a year. In the low cognitive load condition, participants were presented with the employee personality information first, then they rated the likability of the employee and their desire to supervise the employee in a real-world workplace. These questions served as manipulation checks for the personality vignettes and encouraged the participants to quickly form a general person impression or prototype of the fictional employee. Participants were then shown graphs of the employee's sales performance for each quarter on separate web pages so the sales trend would not be immediately obvious. The total annual sales were the same in all conditions ($80,000), but the sales either increased or decreased by $10,000 per quarter depending on the trend condition. After participants viewed the sales information, they were once again provided with the employee personality information and then they rated the employee's sales performance, ability, effort, citizenship, and another manipulation check question about the direction of the sales performance trend.

In the high cognitive load condition, participants were presented with all of the vignettes simultaneously. That is, they first viewed the personality information for all four employees on separate web pages; they rated the likability and desire to supervise all four employees; then they were provided with the quarterly sales performance for all employees (each quarter was on a separate web page); and, finally, they rated each employee's sales performance, effort, ability,

citizenship, and performance trend. The high cognitive load condition was designed to replicate a real-world performance rating context where participants had to process disordered and complex combinations of employee performance information. Differences in the accuracy of the performance trend ratings among the within-subjects conditions indicate that raters may have engaged in more controlled information processing for some employees.

Participants were randomly assigned to the two between-subjects conditions and the order of the vignettes was randomized for each participant. Some participants logged in to the online experiment at approximately the same time, which is why the sample could not be equally distributed between the cognitive load conditions (n = 77 and n = 81 in the low and high cognitive load conditions, respectively). Employee gender was partially counterbalanced by the personality and sales trend conditions such that women were portrayed in the likable-improving trend condition and the dislikable-declining trend condition, whereas men were portrayed in the dislikable-improving trend condition and the likable-declining trend condition.

Personality Vignette Development

I developed vignettes similar to those used by Dobbins and Russell (1986) to manipulate the likeability of the ratee. The vignettes recounted the descriptions of the new employees' behaviors by the more experienced employees and consisted of a number of likable and dislikable personality trait adjectives. To determine which adjectives to include in the vignettes, I consulted Andersen's (1968) study of the likableness of 555 personality trait adjectives. I classified 218 trait adjectives that received high, medium, and low likability ratings into each of the six HEXACO personality domains. The interactions in Study 1 were focused on extraversion, conscientiousness, and agreeableness; therefore, I included two adjectives for each of these traits in the vignettes. The likable employee condition contained trait adjectives that were rated highly

on Andersen's list, whereas the dislikable conditions contained adjectives with lower ratings on the list. Consistent with Dobbins and Russell, I added realism by also including an Honesty-Humility adjective with a moderate likability rating to all the conditions.

One of the likable vignettes was: "One of the new hires is Jane. The first peer you asked described her as friendly, reliable, but somewhat self-centered. Another peer described her as interesting and understanding. You also noticed that Jane is always at work on time and seems to be pleasant when speaking with customers on the phone." A dislikable vignette was: "Another new hire is Bill. The first peer you asked described him as obnoxious and unreliable, but honest. Another peer described him as loud-mouthed and critical. You also noticed that Bill is often late for work and, while he is outgoing when speaking with customers on the phone, he is argumentative at times."

Measures

Manipulation checks. Participants rated the likability of employees with an item similar to that used by Dobbins and Russell (1986). It was revised slightly to enhance the overall realism of the assessment and prime participants to quickly form a general person impression of the fictional employees. The question was, "Think of someone you know who is similar to X. Please rate how much you like that person" and it was rated on a five-point scale (1 = *strongly dislike* to 5 = *strongly like*). Participants also rated their desire to supervise the fictional employees with the following question, "Based on the information you have, how much would you like to supervise X in a real-world workplace?" on a five-point scale (1 = *not at all* to 5 = *very much*). To check the sales trend conditions, participants rated sales trend of each employee on a three-point scale: "How would you describe X's sales performance over the year?" (1 = *declining*, 2 = *consistent*, 3 = *improving*). As described below, this item was recoded to derive a measure of

trend perception accuracy. The amount of time (in seconds) it took participants to provide ratings for each employee was also measured as a cognitive load manipulation check. Presumably, participants who took more time to provide ratings were under more cognitive load than those who took less time.

Performance and attributions. To determine how performance trend and personality information influenced task performance ratings, participants rated the fictional employees' overall sales performance for the entire year. Participants also rated employee ability, effort, and citizenship (getting along and working well with others). Consistent with Study 1, all of the items were rated on a five-point scale (1 = *unsatisfactory* to 5 = *exceptional*).

Performance Trend Accuracy. Raters who accurately perceived the performance trend information likely engaged in controlled information processing, particularly in the high cognitive load condition. Because the performance trend manipulation check item had correct answers depending on the condition (employees had either improving or declining trends), I recoded the item responses to create a measure of performance trend accuracy. To do this, the responses for the vignettes with an improving trend were centered on three (the improving trend rating) and the responses for the declining trend vignettes were centered on one (the declining trend rating). I then recoded the centered scores so that three represented the highest accuracy and one represented the lowest accuracy in both the improving and declining conditions.

Control Variable. Given the potential for gender bias to influence the ratings, participant gender was entered as a control variable in all analyses.

Analysis

Repeated measures ANOVAs were conducted to test the within- and between-subjects main effects for the manipulation check and performance rating items. Unless otherwise

specified, the lower bound degrees of freedom adjustment was used to test the statistical significance of the *F* values for the within-subjects conditions because the sphericity assumption was violated in most of the analyses (i.e., Mauchley's test was statistically significant). To understand whether ratee likability or category inconsistency (i.e., incongruent personality and performance trends) caused participants to engage in more controlled information processing, I also tested the interaction between the within-subjects and between-subjects conditions on the performance trend accuracy variable. Finally, I conducted mediation analysis to determine if trend perception accuracy, which is indicative of controlled information processing, mediated the relationships between either employee likability or category inconsistency with ratings of overall sales performance. Four within-subjects responses were clustered by each participant, so I conducted a multilevel model and entered the within subjects responses at level 1, while allowing the intercepts to randomly vary at level 2 (the participant level) to account for the nested nature of the data. The statistical significance of the indirect effects were tested with the Delta method (MacKinnon, 2008) in MPlus 7.3 because Preacher and Hayes' (2004) bootstrap resampling procedure cannot be conducted with multilevel models.

STUDY 2 RESULTS

Manipulation Checks

There was a within-subjects main effect for ratings of employee likability, F (1, 154) = 186.62, $p < .001$, and desire to supervise, F (1, 154) = 134.80, $p < .001$. As shown in Table 5, the means in the likable conditions were higher than the means in the dislikable conditions for both variables. Pairwise comparisons with the Bonferroni adjustment showed that the differences between the likable and dislikable conditions were statistically significant (all differences

significant at $p < .001$), but the differences between the means in the two likable conditions and two dislikable conditions were not significant.

As shown in Table 5, there was also a within-subjects main effect for the performance trend rating, F (1, 154) = 211.98, $p < .001$. Pairwise comparisons with the Bonferroni correction showed statistically significant differences between vignettes with improving trends and vignettes with declining trends (all differences were significant a $p < .001$). The means in the two improving conditions did not differ from each other, nor did the means in the two declining conditions. The results suggest that participants appropriately interpreted the performance trend information provided in the vignettes.

The checks for cognitive load also demonstrated that this manipulation had the desired effects. Time to conduct the evaluations was highly skewed in both conditions so it was log transformed to normalize the distribution. The between-subjects main effect was statistically significant, F (1, 154) = 6.66, $p = .011$, as participants in the high load condition generally took longer to complete the ratings ($M = 3.39$ and $M = 2.53$ for the high and low load conditions, respectively).

Within-Subjects Main Effects

I then tested the main effects of the within-subjects conditions on ratings of sales performance, effort, ability, and citizenship to compare the results with previous experimental studies and the results of Study 1. There was a main effect for ratings of sales performance, F (1, 154) = 55.32, $p < .001$, as the two conditions with improving performance trends received higher ratings than the conditions with declining trends (see Table 5). Pairwise comparisons with the Bonferroni correction showed that the differences between the improving and declining trends

conditions were statistically significant (all differences significant at $p < .001$), but neither the ratings within the improving nor the declining trend conditions were different from each other.

There was a similar main effect for ratings of employee ability, F (1, 154) = 45.42, $p < .001$; however, the pattern of effects was slightly different than the sales performance evaluations. Table 5 shows that the highest average ratings were in the likable-improving condition, followed by the dislikable-improving condition, then the likable-declining, and the dislikable-declining condition. Pairwise comparisons showed that all of the mean differences were statistically significant (all p-values < .003).

There was a main effect for the effort ratings, F (1, 154) = 63.41, $p < .001$, and Table 5 shows that the pattern of effects was similar to the ratings of employee ability. All of the mean differences were statistically significant except for the differences between the second and third condition. Finally, there was also a main effect for the employee citizenship ratings, F (1, 154) = 153.84, $p < .001$, but the only statistically significant mean differences were between the likable and dislikable conditions (p-values < .001).

Interaction Effect and Mediation Models

Next, I tested the interaction between the within- and between-subjects conditions on performance trend accuracy to determine whether ratee likability or personality-trend inconsistency caused controlled information processing. The interaction effect was statistically significant, F (1, 154) = 6.48, $p = .012$, and the results displayed in Figure 3 show clear differences in accuracy between the cognitive load conditions. The ratings were generally highly accurate in the low cognitive load condition and markedly less accurate in the high load condition; however, there were accuracy differences between vignettes in the high load condition. Paired samples t-tests were conducted to follow-up the mean differences in the high

cognitive load conditions. The results showed that there were significant differences in accuracy ratings for the vignette with a dislikable personality and a declining trend (condition four) compared to conditions two (dislikable personality-improving trend) and three (likable personality-declining trend: t (80) = 3.67, $p < .001$ and t (80) = 3.30, $p = .001$ for comparisons with conditions two and three, respectively). There was also a significant difference between condition one (likable-improving trend) and condition two (dislikable-improving trend), t (80) = 3.67, $p = .021$, while the mean difference between conditions one and three was only marginally significant, t (80) = 1.86, $p = .066$. As the ratings were generally more accurate in conditions two and three, these results indicate that incongruence between personality and performance trends, rather than employee likability, caused raters to engage in controlled information processing.

I conducted mediation analysis to further test the competing explanations. Dummy variables were created for the likability conditions (dislikable vs. likable) and personality-trend consistency conditions (consistent vs. inconsistent). The first model tested if the relationship between employee likability and the overall sales performance rating was mediated by performance trend accuracy. As shown in Figure 4, the relationship between likability and accuracy was not statistically significant, nor was the indirect effect of likability on the sales performance rating. Figure 4 also shows that the effect of personality-trend inconsistency on rating accuracy was significant as was the indirect effect on sales performance ratings.

Taken together, the results of Study 2 suggest that raters relied on performance trend information to derive the sales performance ratings and personality information to derive the citizenship ratings, which are consistent with the results of Study 1. It appears that raters used both performance trend and personality information to derive the ability and effort ratings.

Finally, personality-performance trend inconsistency, rather than ratee likability, appeared to cause raters to engage in controlled information processing.

GENERAL DISCUSSION

Although the results are nuanced, the main findings provide some clarification about how raters integrate performance trend and personality information to derive performance judgments. First, Study 1 showed that raters in field settings indeed appear to observe and apply trend information to derive overall evaluations, which replicates findings from previous experimental research (DeNisi & Stevens, 1981; Lee & Dalal, 2011; Reb & Cropanzano, 2007; Reb & Greguras, 2010; Reb & Cropanzano, 2007; Rudolph et al., 2015). Second, and contrary to previous research, raters did not appear to use trend information to make attributions about ratee characteristics that were peripheral to the trend domain. In both studies, task performance trends only predicted task ratings, whereas personality only predicted behavioral criteria such as coachability and citizenship. Finally, interactions between personality and behavioral trends had not been tested previously, and the effects observed in Studies 1 and 2 suggest that raters engage in controlled information processing about trend information when performance trends are incongruent with personality. Below I discuss how these results can refine research and practice as well as contribute to theory development about the role of memory and cognition in the performance judgment process.

The first implication of the research is that raters in field settings derive task performance judgments based, in part, on performance trend information – raters naturally store and retrieve longitudinal trend information to make performance judgments. These findings appear to represent more than a straightforward replication and extension of prior experimental studies and they have important theoretical implications. In some cases, it appears that the “general

evaluative concept" or prototype that raters used to classify ratees may not have had undue influence on the final ratings. That is, raters appeared to process personality and performance trend information as it unfolded and updated their evaluations accordingly. This may be more consistent with a script schema, rather than a prototype or cognitive categorization schema, and it challenges assumptions that raters apply simple heuristics that often produce biased evaluations.

Although raters certainly appear capable of engaging in controlled information processing of performance trend information, Study 2 showed that this primarily occurs when performance trends are incongruent with general evaluative concepts that were initially derived from observing the ratee's behavior. Moreover, the interaction effect observed in Study 1 suggested that personality-trend inconsistency only led to deeper information processing for targets with less likable traits. Perhaps "bad is stronger than good" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001) and raters have more intense feelings toward people they dislike as compared to likable subordinates. Thus, they may be more "surprised" when dislikable subordinates improve their performance, suggesting that category inconsistent information is more salient for dislikable as compared to likeable subordinates. Future research should explore this possibility and theories of social cognition may require revision to precisely account for the types of situations and people that lead to these different forms of social information processing.

The second implication is that, in Study 1, the coaches did not attribute trait-relevant behaviors to targets based on task performance trends. Instead, raters appeared to apply information about targets' personality traits to derive their appraisals for the behavioral performance dimensions. These results appear to contradict findings from earlier experimental research (DeNisi & Stevens, 1981; Reb & Greguras, 2010; Rudolph et al., 2015), which showed that raters attributed greater effort and ability to ratees based on dynamic trend information. This

is important because it may suggest raters in field settings apply both prototype and script schemas to process social and task performance information – simple cognitive categorization heuristics may have less influence on performance judgments than indicated by previous experimental research.

The results of Study 2 fall somewhere in between Study 1 and prior experimental studies. Consistent with Study 1, raters only used personality information to derive citizenship ratings and performance trend information to derive task performance ratings. However, participants appeared to apply both personality and performance trend information to make attributions about the targets' ability and effort. This apparent inconsistency may be a consequence of the demand characteristics of lab studies: participants can only derive ratings from information presented to them. In the field, supervisors and subordinates are usually involved in a long-term relationship and supervisors have a much larger and richer sample of behavioral information to draw upon when evaluating performance. Unlike participants in experimental research, raters in field settings have no need to attribute trait-relevant behaviors to subordinates based on performance trends because they can observe the person's behavior directly (work ethic, locus of control, interactions with others, etc.).

The effects observed in the current research may have important implications for future theorizing. In particular, it appears that raters apply both person-prototype and script information processing schemas to derive performance judgments. The person-prototypes seemed to dominate ratings of behavioral performance criteria (e.g., citizenship, coachability). When it came to task performance, raters seemed to apply prototype schemas first and only think in terms of scripts (i.e., did the ratee's performance follow the expected pattern over time?) when there were obvious inconsistencies between general impressions of ratees and dynamic performance.

While past research showed that the type of information processing schema could be caused by task instructions (Foti & Lord, 1987), these findings show that ratee characteristics induce different schemas. Although inconsistent with previous research on performance trend processing heuristics, the findings support Srull and Wyer's (1989) model of social cognition – judges reconsider general impressions when faced with enough contradictory information.

Expertise and cognitive load could explain some of the differences between the Study 1 and 2 results and also provide insight for future theory development. The coaches in Study 1 appeared to be using a non-compensatory decision-making heuristic in that they carefully considered performance trends only when players demonstrated more "negative" personality traits, whereas the raters in Study 2 appeared to superficially consider both their general impressions of the ratee and objective performance before deliberating more carefully if discrepancies were observed. Luan and Reb (2017) showed that expert judges are more likely to use non-compensatory heuristics and the coaches may well have had more experience evaluating performance and better understanding of the people they were rating than the participants in Study 2. Cognitive load also likely influences the use of information processing schemas and decision-making heuristics. In the low cognitive load condition of Study 2, all sources of performance information were accurately reflected in the overall performance ratings, whereas participants clearly relied on information processing schemas to derive some of the ratings in the high load condition. Cognitive load and the resulting use of schemas in Study 1 was likely even higher than the manipulated load condition in the second study given that coaches were rating more people and had to recall information over a longer period of time.

To summarize, the results produce at least three propositions that could be integrated into theories of social cognition and performance appraisal. First, raters are capable of applying both

prototype and script information processing schemas and the type of schema applied depends on the characteristics of the task as well as inconsistencies between ratee personality and the expected performance trend. Second, expertise may influence how raters consider different performance dimensions and more expertise may lead to non-compensatory decision-making heuristics. Finally, cognitive load affects the extent to which raters rely on information processing schemas.

Limitations

As with all research, this study is not without limitations. The first is that the findings may not be generalizable to the workplace. For some positions, managers do not have access to longitudinal, objective performance information, which would limit their ability to apply this information to performance evaluations. Moreover, specific performance episodes may not be as closely scrutinized in the workplace as in the athletic context or an experimental study, where performance information is carefully reviewed and performance trends are quite salient. However, there are numerous contexts where objective, longitudinal performance data is available (e.g., sales, manufacturing, software development), and longitudinal trend information could be incorporated into evaluations. Even for the contexts without clear objective numbers, raters evaluate performance somehow, and likely have a performance trend in mind, however informal.

There may also be potential endogeneity concerns in Study 1. It is possible that the best coaches helped players improve more over the course of the season and had a tendency to provide higher appraisals, or raters with certain personality traits have rating tendencies. To account for these potential problems, dummy variables for team and position group were controlled when testing the research questions in Study 1. This approach accounts for the time

invariant influences of unobserved team and position-level effects, including the systematic rating tendencies of the coaches. Future research should seek to further measure these effects and determine how group characteristics or tendencies of evaluators influence how longitudinal performance information is incorporated into overall evaluations.

Finally, and as mentioned previously, Study 2 may lack external validity. It is possible that raters process information differently when evaluating fictional employees as compared to real people. Unlike this research, previous experimental studies did not include personality information in the employee vignettes and there were some differences in the results of Study 2 and prior experimental research. Researchers are encouraged to further develop experimental protocols to create high fidelity simulations of the real-world performance rating process.

Practical Implications

This research has a number of practical implications for the design and implementation of performance evaluation systems. It is commendable for raters to recognize the efforts of ratees, particularly when that effort does not result in observable performance outcomes, or when the situation may not allow for marked improvement in objective performance. However, the results suggest that personality and interpersonal interactions can affect a rater's perception of performance trends, particularly when there is relative congruence between personality and trends. In fact, the results of Study 2 suggested that sales performance rating of the likable employee with an improving trend was downwardly biased because the raters perceived the trend less accurately than trends of employees with inconsistent characteristics. Raters must be aware of how their general impressions may cause them to disregard important performance information, especially for those who display more prosocial traits. To avoid such biases,

managers may need to explicitly consider objective performance trends throughout the evaluation period.

Although fraught with challenges, supervisor performance appraisals are a foundational component of HR management systems and are often the primary criterion variable in many programs of research. The results of this study help clarify the processes by which raters render overall evaluations, and indicate that managers should explicitly consider the role of personality and dynamic performance trends when evaluating subordinate performance.

APPENDIX: STUDY 2 PILOT TESTING PROCEDURES AND RESULTS

Pilot Test 1: Vignette Development

Sample

To understand how people with employment experience process and evaluate performance information, 20 participants were recruited from Probit, a Canadian online survey panel. Most of the participants were employed full-time (48.8%) or semi-retired (29.3%), and 53.7% of the participants were male.

Procedure

Participants read introductory information asking them to assume the role of a sales manager who was required to evaluate the performance of two new salespeople. Participants were then provided with personality and performance trend information for the two employees, which were presented in random order for each participant. More detailed descriptions of introductory information and the employee vignettes are provided in the Study 2 Procedure section.

Measures

Participants rated three manipulation check items: likableness of the fictional employees, their desire to supervise the employees in a real-world workplace, and the direction of the performance trend. They also rated four performance dimensions, which included overall sales, effort, ability, and citizenship. The sales trend direction was rated on a three-point scale and the remainder of the items were rated on a five-point scale. More detailed descriptions of the items and the rationale for including them are provided in the Study 2 Measures section.

Results

I conducted paired samples t-tests to compare the means between the two within-subjects conditions. There were significant differences in the expected direction for all of the items (Table A1). There was no variance in responses for the sales trend manipulation check and a *t-test* could not be conducted.

Table A1: t-test Results

Item	Condition 1 (likable - improving trend)		Condition 2 (dislikable - declining trend)		
	M	*SD*	*M*	*SD*	*t value*
Manipulation Checks					
Likability	3.70	0.66	2.45	0.83	4.80**
Desire to supervise	3.50	0.83	1.95	1.00	5.11**
Sales trend	3.00	0.00	1.00	0.00	--
Performance Dimensions					
Sales performance	3.75	0.64	2.05	0.76	6.74**
Ability	3.85	0.67	2.50	1.00	4.61**
Effort	3.95	0.76	2.25	0.64	6.74**
Citizenship	3.65	0.49	1.90	0.45	10.93**

** $p < .01$

Pilot Test 2: Cognitive Load Manipulation

The second pilot was designed to test the effects of rater cognitive load during the performance evaluation. I used the two personality vignettes from the first pilot test and developed two additional vignettes (one likable and one dislikable), so this pilot study contained the four within-subjects (likable-improving trend, dislikable improving trend, likable-declining trend, dislikable-declining trend) and two between-subjects conditions (low vs. high cognitive load).

Based on the Study 1 results, I expected two primary effects. First, the ratings in the low cognitive load condition would be "more accurate" than in the high cognitive load condition and participants would take longer to complete the evaluations in the high load condition. Second, the high cognitive load condition was designed to replicate a real-world performance rating context; as such, I expected to see some evidence of a similar

interaction effect to that observed in Study 1. That is, performance ratings in the high load condition would be more accurate from some employees than others, indicating that certain characteristics of the employees caused participants to engage in more controlled (deliberative) information processing.

Sample

The sample consisted of 23 participants recruited from Probit. A slight majority of the participants were female (52.2%). Much of the sample were middle-aged (47.6% were in the 35-54 age range) and another 26.1% of the sample were 65 or older. The majority of the sample has some higher education as 34.8% had obtained a college or technical diploma and 38.7% had a bachelor's or graduate degree. Most of the participants were employed full-time (43.5%), part-time (13.0%), or were self-employed (17.4%). Twenty-one percent were retired and one of the participants was a student.

Procedure

Participants read the same introductory information as described in Pilot Test 1. In the low cognitive load condition, participants were presented with one vignette at a time. That is, they viewed the personality description for one employee, rated the likability and their desire to supervise the employee, viewed the performance information for the year (each quarter's sales was on a separate web page), and then rated the employee's sales performance, effort, ability, citizenship, and performance trend.

In the high cognitive load condition, participants were presented with all of the vignettes simultaneously. That is, they first viewed the personality information for all four employees on separate web pages; rated the likability and desire to supervise all four employees; then were provided with the quarterly sales performance for all employees

(each quarter was on a separate web page); and they finally rated each employee's sales performance, effort, ability, citizenship, and performance trend. The amount of time (in seconds) participants spent conducting the evaluations was also recorded as a manipulation check for cognitive load.

Results

Repeated measures analysis of variance was conducted to test the within-subjects main effects and interactions with the between-subjects cognitive load condition for the manipulation check measures. For most analyses, the sphericity assumption was violated (i.e., Mauchly's test of sphericity was statistically significant), so I interpreted the *F* tests based with the lower bound degrees of freedom adjustment in these cases.

As shown in Table A2, the repeated measures ANOVA results showed a main effect of employee likability and desire to supervise, with the vignettes designed to be more likable receiving higher overall ratings for both variables. The sphericity assumption was not violated for analysis of the sales trend manipulation check and the results showed a within-subjects main effect, $F(3, 63) = 27.45$, $p < .001$, indicating that respondents appropriately detected the direction of sales trends presented in the vignettes. There was a between-subjects main effect for the evaluation time manipulation check for cognitive load, $F(1, 21) = 5.14$, $p = .034$ (not reported in the table). Participants spent more time evaluating employees in the high vs. low load condition ($M = 54.21$ seconds for the high load condition versus $M = 38.15$ seconds for the low load conditions). There was also an interaction between the within-subjects conditions and the between-subjects cognitive load condition on the performance trend ratings, $F(3, 63) = 4.79$, $p = .005$. All of the ratings in the low cognitive load condition were relatively accurate, whereas participants in the high

load condition provided relatively accurate trend ratings for two employees (i.e., ratings were more similar to those in the low cognitive load condition), but relatively inaccurate ratings for the other two employees (i.e., ratings were dissimilar to the low load condition). This effect suggests that participants engaged in deeper information processing for some employees and may allow for a test of the theoretical mechanisms that were applied to explain the Study 1 results.

Table A2: Means and Standard Deviations for Pilot Study 2 Conditions

		Within-Subjects Condition									
Rating Dimension		Condition 1 (Improving-Likable)		Condition 2 (Improving-Dislikable)		Condition 3 (Declining-Likable)		Condition 4 (Declining-Dislikable)			
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	η^2
Manipulation Checks											
Likability	23	3.78	0.95	2.04	1.02	3.83	0.78	2.39	0.89	22.17**	.51
Desire to Supervise	23	3.70	1.06	2.00	1.09	3.91	0.79	2.48	1.24	17.11**	.45
Performance Trend	23	2.57	0.79	2.78	0.52	1.22	0.52	1.61	0.78	27.45**	.57
Performance Ratings											
Sales Performance	23	3.47	0.67	3.57	0.84	3.00	0.52	2.70	0.76	8.41**	.29
Ability	23	3.65	0.78	3.52	0.79	3.09	0.67	2.87	0.76	6.43**	.23
Effort	23	3.70	0.82	3.43	0.95	3.04	0.37	2.52	0.73	10.03**	.32
Citizenship	23	3.78	0.74	2.22	0.67	3.91	0.79	2.17	0.78	34.89**	.62
Sales Trend											
Low Cognitive Load	12	2.92	0.29	2.75	0.62	1.08	0.29	1.25	0.45	4.79**	.19
High Cognitive Load	11	2.18	0.98	2.81	0.40	1.36	0.67	2.00	0.89		

Note. The *F* test for the Sales Trend variable is for the within- by between-subjects condition interaction. The remainder of the *F* tests are for the within-subjects main effects.

* $p < .05$, ** $p < .01$

Pilot Test 3: Controlled Information Processing Measure

I conducted a third pilot study to test another measure of controlled (deliberative) information processing in addition to the performance trend accuracy. Participants were given the option to review an overall sales summary of each employee prior to conducting the performance ratings. I hypothesized that participants who chose to review the sales summary were seeking confirmatory information to support their evaluations and were engaged in more controlled information processing.

A total of 20 people were recruited from Probit and were randomly assigned to the between-subjects conditions (10 per condition). Unfortunately, the cognitive effort measure did not work as expected. In the low cognitive load condition, the sales summary was viewed 55% of the time, while it had zero variance in the high load condition as it was viewed 100% of the time. This serves as a manipulation check such that participants in the high condition were indeed experiencing increased cognitive load and needed to view the sales summaries; however, because the variable was a constant in the high load condition it could not be used as an information processing measure in subsequent analyses.

TABLE 1

Game Performance Statistics by Position

Games Statistic	Defensive Back	Defensive Line	Linebacker	Quarterback	Receiver	Running Back
Tackles	X	X	X			
Pass Breakups	X					
Interceptions	X					
Tackle-for-loss		X	X			
Completion Percentage				X		
Sacks Taken (reversed)				X		
Passing Yards				X		
Interceptions Thrown (reversed)				X		
Touchdowns Thrown				X		
Points Scored				X	X	X
Rushing Yards				X		X
Reception Yards					X	X

Note. Some statistics were excluded for certain positions because they had a very low base rate on a game-by-game basis and would have created skewed the distributions if included in the composite. These statistics included pass breakups and interceptions for defensive line and linebackers; tackle-for-loss for defensive backs; and rushing yards for receivers.

TABLE 2
Descriptive Statistics and Correlations among Study 1 Variables

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Performance T1	.12	1.23																			
2. Performance T2	.07	1.19	.24**																		
3. Performance T3	.36	1.60	.30**	.11																	
4. Performance T4	.14	1.33	.24**	.08	.22**																
5. Performance T5	.36	1.55	.28**	.01	.19**	.15*															
6. Performance T6	.07	1.52	.26**	.06	.35**	.30**	.23**														
7. Performance T7	.17	1.46	.13	.16*	.31**	.06	.16*	.16*													
8. Performance T8	.15	1.36	.00	.20**	.09	.17	.10	.22**	.34**												
9. Physical Ability	.20	.48	.20**	.14*	.19**	.26**	.08	.14*	.03	.09											
10. Honesty-Humility	3.05	.52	-.09	.10	-.08	-.09	-.25**	-.05	.01	.04	-.09	*(.70)*									
11. Emotionality	2.50	.43	.02	.08	-.06	.13	-.09	.03	-.05	-.07	-.04	-.09	*(.66)*								
12. Extraversion	3.41	.56	-.05	-.04	-.10	.02	.07	-.01	-.04	.02	.05	-.14*	-.06	*(.79)*							
13. Agreeableness	3.12	.59	.06	-.03	.12	.03	-.05	.10	-.05	-.02	-.02	.35**	-.12	-.16*	*(.77)*						
14. Conscientiousness	3.43	.50	-.01	-.05	-.08	-.09	-.01	.05	-.01	.12	.03	.18*	-.08	.09	.07	*(.74)*					
15. Openness	3.45	.56	.08	.15*	-.06	-.12	.04	.05	-.07	-.19**	.05	.15	-.05	.15*	.13	.19**	*(.75)*				
16. Coachability	3.76	.93	-.04	.09	.08	.07	.00	.03	.03	.17*	-.03	.17*	.03	-.27**	.12	.16*	-.11				
17. Athletic Ability	3.85	.76	.14*	.22**	.22**	.10	.13	.13	.19*	.19**	.26**	-.04	.08	-.03	-.03	.09	.07	.13			
18. Game Performance	3.59	.83	.14*	.14*	.28**	.16*	.27**	.23**	.33**	.25**	.04	-.10	.09	-.13	-.09	-.03	-.15*	.48**	.39**		
19. Team Player	4.00	.84	.01	-.01	.14*	.15*	.11	.14*	.09	.13	-.04	.06	.00	-.20**	.15*	.10	-.12	.65**	.12	.46**	
20. Work Ethic	3.89	.87	-.11	-.02	.09	-.07	.06	.13	.07	.13	.07	-.04	.06	-.05	-.06	.22**	.04	.44**	.19**	.38**	.54**

Note. N = 199. The team and position dummy variable are not included in the table. The physical ability variable does not conform to a normal z-score distribution because the z-scores were derived from the full sample, whereas this subsample only contains players who participated in 3 or more games. Cronbach Alphas are reported in the diagonal where appropriate.
* $p < .05$, ** $p < .01$

TABLE 3

Effects of Performance Trends and Personality Traits on Specific Performance Dimensions

Variable	Coachability	Athletic Ability	Game Performance	Team Player	Work Ethic
Latent Growth Intercept	.19 (.18)	.37 (.13)**	.61 (.15)**	.38 (.18)*	-.12 (.19)
Latent Growth Slope	1.75 (1.38)	2.66 (1.26)*	3.90 (1.33)**	2.29 (1.38)†	2.25 (1.65)
Physical Ability	-.04 (.15)	.32 (.12)**	-.06 (.14)	-.14 (.14)	.20 (.14)
Honesty-Humility	.31 (.13)*	.03 (.13)	.00 (.13)	.11 (.15)	-.07 (.15)
Emotionality	.09 (.14)	.14 (.13)	.17 (.13)	.01 (.14)	.17 (.16)
Extraversion	-.40 (.12)**	-.09 (.11)	-.18 (.12)	-.25 (.11)*	-.09 (.11)
Agreeableness	.01 (.13)	-.11 (.12)	-.12 (.11)	.11 (.13)	-.05 (.12)
Conscientiousness	.30 (.14)*	.09 (.12)	-.04 (.13)	.17 (.12)	.38 (.13)**
Openness	-.17 (.14)	.17 (.11)	-.11 (.11)	-.12 (.12)	.08 (.12)
LGCM Variance Components					
Intercept variance	.49(.15)**	.49(.15)**	.49(.15)**	.49(.15)**	.49(.15)**
Slope Variance	.01 (.01)*	.01 (.01)*	.01 (.01)*	.01 (.01)*	.01 (.01)*
Intercept-slope covariance	-.04 (.03)	-.04 (.03)	-.04 (.03)	-.04 (.03)	-.04 (.03)

Note. $N = 199$. LGCM = Latent Growth Curve Model. Unstandardized effects are reported and standard errors are shown in parentheses. Although not reported in the table, the team and position dummy variables were entered as controls in all analyses.
† $p < .10$, * $p < .05$, ** $p < .01$

TABLE 4

Interactions between Personality Traits and Game Performance Slopes on Game Performance Ratings

Variable	Model 1	Model 2	Model 3
Latent Growth Intercept	.62 (.19)**	.64 (.19)**	.63 (.18)**
Latent Growth Slope	4.23 (1.54)**	4.22 (1.67)*	3.62 (1.39)**
Physical Ability	-.06 (.13)	-.09 (.14)	-.09 (.14)
Honesty-Humility	-.04 (.13)	-.01 (.13)	-.01 (.14)
Emotionality	.20 (.13)	.14 (.13)	.17 (.13)
Extraversion	-.20 (.14)	-.19 (.11)†	-.16 (.12)
Agreeableness	-.15 (.11)	-.08 (.12)	-.13 (.11)
Conscientiousness	-.08 (.13)	-.03 (.12)	-.07 (.15)
Openness	-.10 (.11)	-.07 (.11)	-.10 (.11)
Intercept x Physical Ability	--	--	--
Slope x Physical Ability	--	--	--
Intercept x Extraversion	.17 (.27)	--	--
Slope x Extraversion	3.87 (1.53)*	--	--
Intercept x Agreeableness	--	-.42 (.23)†	--
Slope x Agreeableness	--	-3.42 (1.49)*	--
Intercept x Conscientiousness	--	--	.28 (.33)
Slope x Conscientiousness	--	--	-.67 (2.01)
LGCM Variance Components			
Intercept variance	.49 (.15)**	.49 (.15)**	.50 (.15)**
Slope Variance	.01 (.01)*	.01 (.01)†	.01 (.01)*
Intercept-slope covariance	-.04 (.03)	-.04 (.03)	-.04 (.03)

Note. N = 199. LGCM = Latent Growth Curve Model. Unstandardized effects are reported and standard errors are shown in parentheses. Although not reported in the table, the team and position dummy variables were entered as controls in all analyses.
† $p < .10$, * $p < .05$, ** $p < .01$

TABLE 5

Mean Ratings of the Manipulation Checks, Performance Dimensions, and Sales Trend Accuracy

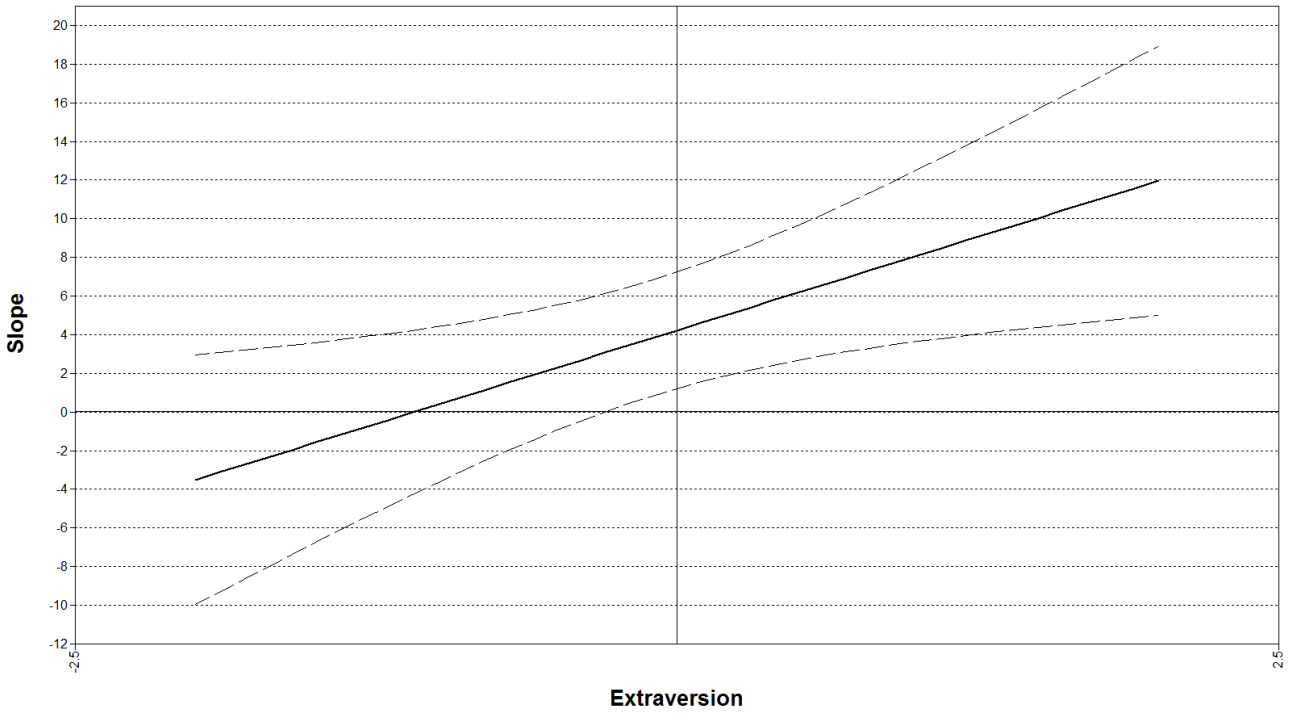
		Within-Subjects Condition									
Rating Dimension		Condition 1 (Improving-Likable)		Condition 2 (Improving-Dislikable)		Condition 3 (Declining-Likable)		Condition 4 (Declining-Dislikable)			
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	η^2
Manipulation Checks											
Likability	158	3.66	0.84	2.09	0.88	3.80	0.96	2.25	0.87	186.62**	.55
Desire to Supervise	158	3.74	0.93	2.13	1.19	3.86	0.99	2.24	1.13	134.80**	.46
Performance Trend	158	2.65	0.64	2.76	0.52	1.25	0.54	1.46	0.73	211.98**	.58
Performance Ratings											
Sales Performance	158	3.50	0.69	3.41	0.68	2.75	0.72	2.67	0.71	55.32**	.26
Ability	158	3.54	0.64	3.28	0.70	2.92	0.74	2.68	0.70	45.42**	.23
Effort	158	3.63	0.65	3.22	0.77	2.99	0.74	2.49	0.76	63.41**	.29
Citizenship	158	3.53	0.64	2.28	0.75	3.53	0.67	2.30	0.74	153.84**	.50
Sales Trend Accuracy											
Low Cognitive Load	77	2.96	0.19	2.94	0.27	2.96	0.19	2.95	0.22	6.48*	.04
High Cognitive Load	81	2.35	0.76	2.59	0.63	2.54	0.67	2.16	0.83		

Note. Participant gender was entered as a control in all analyses. The *F* test for the Sales Trend Accuracy variable is for the within- by between-subjects condition interaction. The remainder of the *F* tests are for the within-subjects main effects.

* $p < .05$, ** $p < .01$

FIGURE 1

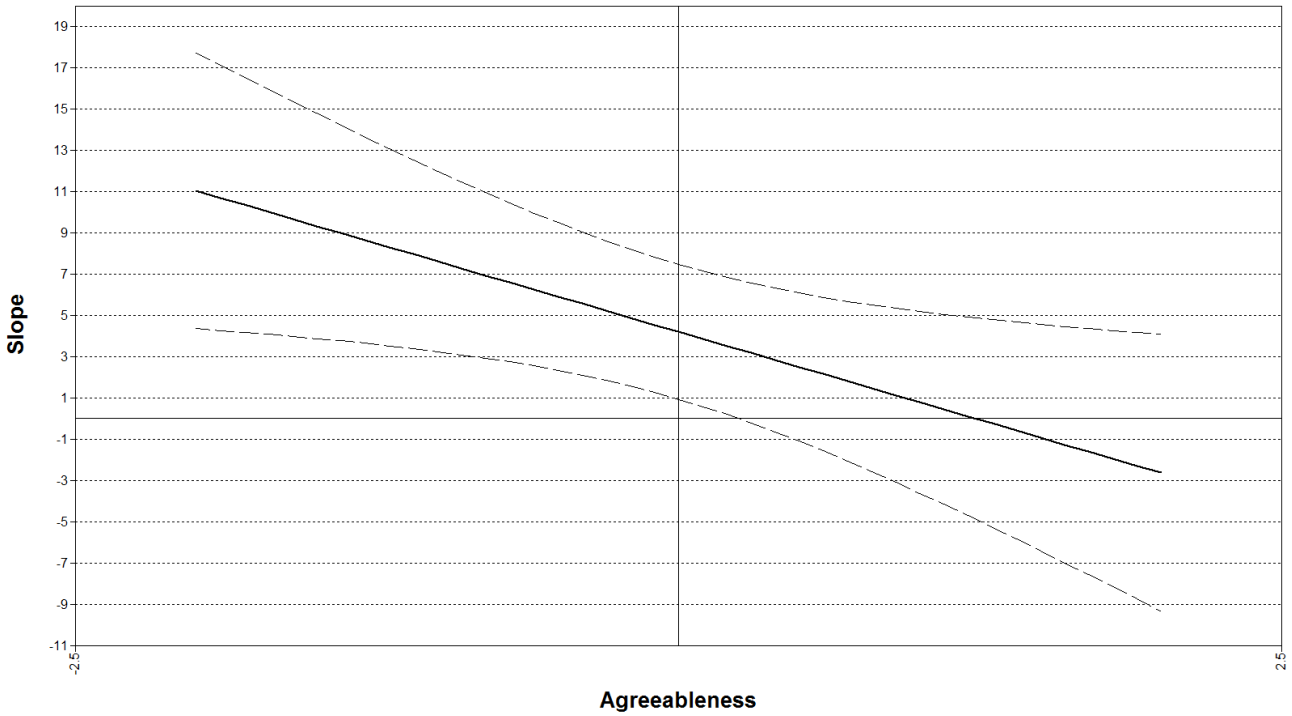
Regions of Significance Plot for the Interaction between the Latent Slope Term and Extraversion on Game Performance Ratings



Note. The solid line represents that value of the unstandardized association between the latent performance slope term and game performance ratings (y-axis) at each level of Extraversion (x-axis). The dashed lines are the upper and lower bounds of the 95% confidence interval. The association between the latent performance slope and performance appraisals is statistically significant for those higher in extraversion (the confidence interval excludes zero) and not significant for those lower in extraversion.

FIGURE 2

Regions of Significance Plot for the Interaction between the Latent Slope Term and Agreeableness on Game Performance Ratings



Note. The solid line represents that value of the unstandardized association between the latent performance slope term and game performance ratings (y-axis) at each level of agreeableness (x-axis). The dashed lines are the upper and lower bounds of the 95% confidence interval. The association between the latent performance slope and performance ratings is statistically significant for those lower in agreeableness (the confidence interval excludes zero) and not significant for those higher in agreeableness.

FIGURE 3

Interaction between Within-Subjects Conditions and Cognitive Load on Accuracy of Sales Trend Ratings

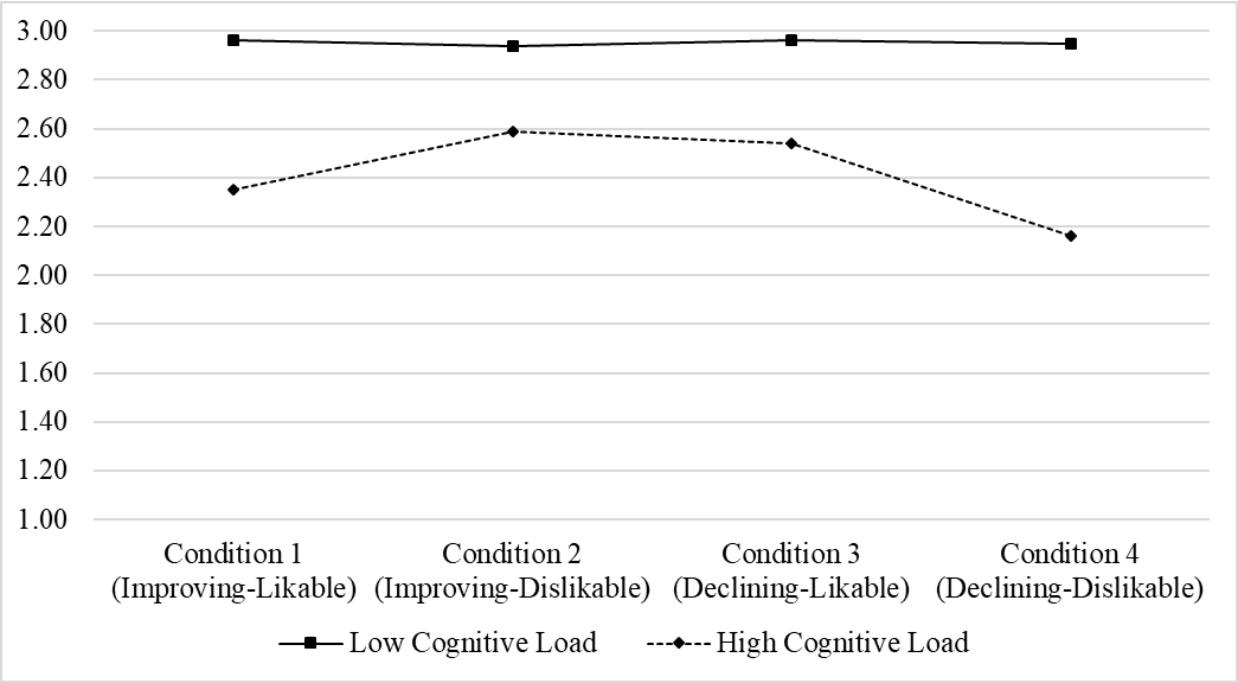
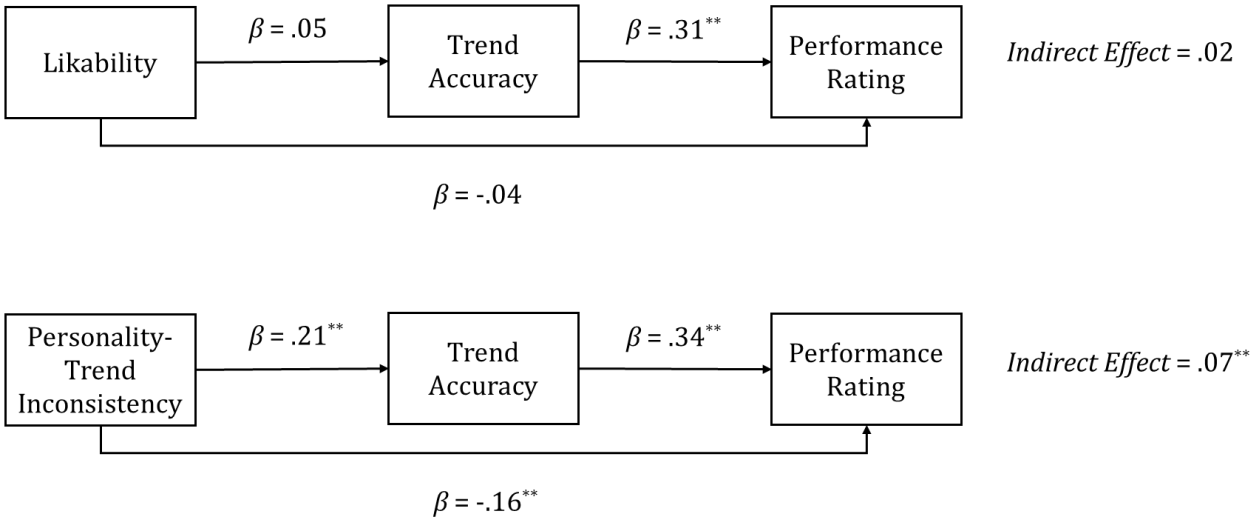


FIGURE 4

Study 2 Indirect Effects



Note. Participant gender was entered as a control in all analyses.
$^{*}p < .05$, $^{**}p < .01$

REFERENCES

Anderson, N. H. 1968. Likableness ratings of 555 personality-trait words. ***Journal of Personality and Social Psychology***, 9: 272-279. *doi: 10.1037/h0025907*

Ashton, M. C., & Lee, K. 2009. The HEXACO-60: A short measure of the major dimensions of personality. ***Journal of Personality Assessment***, 91: 340–345. *doi: 10.1080/00223890902935878*

Barnes, C. M., & Morgeson, F. P. 2007. Typical performance, maximal performance, and performance variability: Expanding our understanding of how organizations value performance. ***Human Performance***, 20: 259–274. *doi: 10.1080/08959280701333289*

Barnes, C. M., Reb, J., & Ang, D. 2012. More than just the mean: Moving to a dynamic view of performance-based compensation. ***Journal of Applied Psychology,*** 97: 711–718. *doi: 10.1037/a0026927*

Barrick, M. R., Mount, M. K., & Judge, T. A. 2001. Personality and performance at the beginning of the new millennium: What do we know and where do we go next? ***International Journal of Selection and Assessment***, 9: 9-29. *doi: 10.1111/1468-2389.00160*

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. 2001. Bad is stronger than good. ***Review of General Psychology***, 5: 323-370. *doi: 10.1037/1089-2680.5.4.323*

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. 1995. On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. ***Personnel Psychology***, 48: 587–605. *doi: 10.1111/j.1744-6570.1995.tb01772.x*

Byrne, B. M. 2012. ***Structural equation modeling with MPlus: Basic concepts, applications, and programming.*** New York, NY: Taylor & Francis.

Day, D. V, Gordon, S., & Fink, C. 2012. The sporting life: Exploring organizations through the lens of sport. ***Academy of Management Annals***, 6: 397-433. *doi: 10.1080/19416520.2012.678697*

Day, D. V., Sin, H. P., & Chen, T. T. 2004. Assessing the burdens of leadership: Effects of formal leadership roles on individual performance over time. ***Personnel Psychology***, 57: 573-605. *doi: 10.1111/j.1744-6570.2004.00001.x*

DeNisi, A. S., Cafferty, T., & Meglino, B. 1984. A cognitive view of the performance appraisal process: A model and research implications. ***Organizational Behavior and Human Performance,*** 33: 360-396. *doi: 10.1016/0030-5073(84)90029-1*

DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? ***Journal of Applied Psychology***, 102: 421-433. *doi: 10.1037/apl0000085*

DeNisi, A. S., & Pritchard, R. D. 2006. Performance appraisal, performance management and improving individual performance: A motivational framework. ***Management and Organization Review***, 2: 253-277. *doi: 10.1111/j.1740-8784.2006.00042.x*

DeNisi, A. S., & Sonesh, S. 2011. The appraisal and management of performance at work. In S Zedeck (Ed.) ***APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization*****:** 255-279. Washington, DC: American Psychological Association.

DeNisi, A. S., & Stevens, G. E. 1981. Profiles of performance, performance evaluations, and personnel decisions. ***Academy of Management Journal***, 24: 592-602. *doi: 10.2307/255577*

Dobbins, G. H., & Russell, J. M. 1986. The biasing effects of subordinate likeableness on leaders' responses to poor performers: A laboratory and a field study. ***Personnel Psychology***, 39: 759–777. *doi: 10.1111/j.1744-6570.1986.tb00593.x*

Erez, A., Schilpzand, P., Leavitt, K., Woolum, A. H., & Judge, T. A. 2015. Inherently relational: Interactions between peers' and individuals' personalities impact reward giving and appraisal of individual performance. ***Academy of Management Journal***, 58: 1761–1784. *doi: 10.5465/amj.2011.0214*

Favero, J. L., & Ilgen, D. R. 1989. The effects of ratee prototypicality on rater observation and accuracy. ***Journal of Applied Social Psychology***, 19: 932–946. *doi: 10.1111/j.1559-1816.1989.tb01230.x*

Fisher, C. D. 2008. What if we took within-person performance variability seriously? ***Industrial and Organizational Psychology,*** 1: 185–189. *doi: 10.1111/j.1754-9434.2008.00036.x*

Foti, R. J., & Lord, R. G. 1987. Prototypes and scripts: The effects of alternative methods of processing information on rating accuracy. ***Organizational Behavior and Human Decision Processes,*** 39: 318–340. *doi: 10.1016/0749-5978(87)90027-6*

Harris, M. 1994. Rater motivation in the performance appraisal context: A theoretical framework. ***Journal of Management***, 20: 737−756. *doi: 10.1016/0149-2063(94)90028-0*

Harris, M. M., Ispas, D., & Schmidt, G. F. 2008. Inaccurate performance ratings are a reflection of larger organizational issues. ***Industrial and Organizational Psychology*****,** 1: 190–193. *doi: 10.1111/j.1754-9434.2008.00037.x*

Hofmann, D. A., Jacobs, R., & Baratta, J. E. 1993. Dynamic criteria and the measurement of change. ***Journal of Applied Psychology***, 78: 194-204. *doi: 10.1037/0021-9010.78.2.194*

Hofmann, D. A., Jacobs, R., & Gerras, S. J. 1992. Mapping individual performance over time. ***Journal of Applied Psychology,*** 77: 185-195. *doi: 10.1037/0021-9010.77.2.185*

Hogan, J., & Holland, B. 2003. Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. ***Journal of Applied Psychology***, 88: 100–112. *doi: 10.1037/0021-9010.88.1.100*

Ilgen, D. R., 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. *doi: 10.1006/obhd.1993.1015*

Ilgen, D. R., & Feldman, J. M. 1983. Performance appraisal: A process focus. In L. Cummings & B. Staw (Eds.), ***Research in organizational behavior***: 141-197. Greenwich, CT: JAI Press.

Lee, K. & Ashton, M. C. n.d. ***The HEXACO Personality Inventory – Revised Scale Descriptions.*** Retireived from www.hexaco.org/scaledescriptions.

Lee, H., & Dalal, R. S. 2011. The effects of performance extremities on ratings of dynamic performance. ***Human Performance***, 24: 99–118. *doi: 10.1080/08959285.2011.554467*

Luan, S., & Reb, J. 2017. Fast-and-frugal trees as noncompensatory models of performance-based personnel decisions. ***Organizational Behavior and Human Decision Processes***, 141: 29–42. *doi: 10.1016/j.obhdp.2017.05.003*

Lyons, B. D., Hoffman, B. J., & Michel, J. W. 2009. Not much more than g? An examination of the impact of intelligence on NFL performance. ***Human Performance,*** 22: 225-245. *doi: 10.1080/08959280902970401*

MacKinnon, D. P. 2008. ***Introduction to Statistical Mediation Analysis***. Mahwah, NJ: Erlbaum.

Major, B. 1980. Information acquisition and attribution processes. ***Journal of Personality and Social Psychology***, 39: 1010-1023. *doi: 10.1037/h0077725*

Murphy, K. R., & Cleveland, J. N. 1995. ***Understanding performance appraisal: Social, organizational, and goal-based perspectives***. Thousand Oaks, CA: Sage.

Nisbett, R. E., & Wilson, T. D. 1977. The halo effect: Evidence for unconscious alteration of judgments. ***Journal of Personality and Social Psychology***, 35: 250–256. *doi: 10.1037/0022-3514.35.4.250*

Nisen, M. 2015. Why GE had to kill its annual performance reviews after more than three decades. ***Quartz,*** August 13. Retrieved from qz.com/428813/ge-performance-review-strategy-shift/.

Osterman, P. 2007. Comment on Le, Oh, Shaffer and Schmidt. ***Academy of Management Perspectives,*** 3: 16–18. *doi: 10.5465/AMP.2007.26421234*

Penney, L. M., David, E., & Witt, L. 2011. A review of personality and performance: Identifying boundaries, contingencies, and future research directions. ***Human Resource Management Review,*** 21: 297-310. *doi: 10.1016/j.hrmr.2010.10.005*

Piedmont, R. L., Hill, D. C., & Blanco, S. 1999. Predicting athletic performance using the five-factor model of personality. ***Personality and Individual Differences,*** 27: 769–777. *doi: 10.1016/S0191-8869(98)00280-3*

Ployhart, R. E., & Hakel, M. D. 1998. The substantive nature of performance variability: Predicting interindividual differences in intraindividual performance. ***Personnel Psychology,*** 51: 859–901. *doi: 10.1111/j.1744-6570.1998.tb00744.x*

Preacher, K. J., & Hayes, A. F. 2004. SPSS and SAS procedures for estimating indirect effects in simple mediation models. ***Behavior Research Methods***, 36: 717-731. *doi: 10.3758/BF03206553*

Rabbitt, P., Osman, P., Moore, B., & Stollery, B. 2001. There are stable individual differences in performance variability, both from moment to moment and from day to day. ***The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology,*** 54A(4): 981–1003. *doi: 10.1080/02724980042000534*

Reb, J., & Cropanzano, R. 2007. Evaluating dynamic performance: The influence of salient Gestalt characteristics on performance ratings. ***Journal of Applied Psychology***,92: 490-499. *doi: 10.1037/0021-9010.92.2.490*

Reb, J., & Greguras, G. J. 2008. Dynamic performance and the performance–performance rating relation. ***Industrial and Organizational Psychology***, 1: 194–196. *doi: 10.1111/j.1754-9434.2008.00038.x*

Reb, J., & Greguras, G. J. 2010. Understanding performance ratings: Dynamic performance, attributions, and rating purpose. ***Journal of Applied Psychology***, 95: 213–220. *doi: 10.1037/a0017237*

Rock, D., & Jones, B. 2015. Why more and more companies are ditching performance ratings. ***Harvard Business Review,*** September 8: 2-4.

Rudolph, C. W., Harari, M. B., & Nieminen, L. R. G. 2015. The effect of performance trend on performance ratings occurs through observer attributions, but depends on performance variability. ***Journal of Applied Social Psychology***, 45: 541–560. *doi: 10.1111/jasp.12318*

Scott, W. E., & Hamner, W. C. 1975. The influence of variations in performance profiles on the performance evaluation process: An examination of the validity of the criterion. ***Organizational Behavior and Human Performance***, 14: 360–370. *doi: 10.1016/0030-5073(75)90036-7*

Spence, J. R., & Keeping, L. M. 2011. Conscious rating distortion in performance appraisal: A review, commentary, and proposed framework for research. ***Human Resource Management Review,*** 21: 85–95. *doi: 10.1016/j.hrmr.2010.09.013*

Srull, T. K., & Wyer, R. S. 1989. Person memory and judgment. ***Psychological Review***, 96: 58-83. *doi: 10.1037/0033-295X.96.1.58*

Taylor, M. S., Tracy, K. B., Renard, M. B., Harrison, J. K., & Carroll, S. J. 1995. Due process in performance appraisal: A quasi-experiment in procedural justice. ***Administrative Science Quarterly***, 40: 495–523. *doi: 10.2307/2393795*

Biography

Joseph Schmidt (jschmidt@edwards.usask.ca)

Dr. Schmidt's research is focused on understanding how strategy, human resource management, and leadership interact to influence employee behaviors and firm performance. He also studies group effectiveness and the multilevel influences of group processes on individual and team outcomes. Prior to his academic career he worked for a human resource consulting firm in Vancouver, British Columbia, Canada.