Reconsidering the Accuracy of Follower Leadership Ratings

Abstract

Accurate behavioral measurement is essential to developing a science of leadership, yet accurate measurement has remained elusive. The use of follower reports of leader behavior creates challenges given that a large body of basic and applied research suggests that behavioral ratings reflect not only recall of actual behaviors, but also inferences based on semantic memory, which may vary among individuals. In this paper, we examine several explanations for rater effects that are associated with follower individual differences, contextual factors, and even research methods, such as the type of measure used, that may bias ratings of leader behavior. We also develop a conceptual model to illustrate these processes. Finally, we offer potential solutions to increase accuracy in follower reports of leader behavior.

The leadership field relies heavily on follower ratings of leader behavior both in research settings to test leadership theories and in applied settings for leadership development purposes (e.g., 360-degree feedback). Given that leadership reflects a dynamic interaction between leaders and followers (Riggio, Chaleff, & Lipman-Blumen, 2008; Shamir, Pillai, Bligh, & Uhl-Bien, 2006), follower ratings provide an important source of data. Indeed, Shamir (2007) describes followers as "co-producers of leadership." However, reliance on follower ratings of leader behavior as a key measure of leadership processes, or even as the sole measure, creates significant challenges. In particular, followers become important contributors to the processes they are used to measure, raising both the issue of accuracy of leader ratings and the potential for biases in ratings processes that are associated with individual differences among followers (Bono, Hooper, & Yoon, 2012; Hunter, Bedell-Avers, & Mumford, 2007).

If leader ratings are to be accurate at the behavioral level, they must accurately report whether specific types of behaviors occurred (e.g., Lord, 1985). In terms of signal detection theory, accurate behavioral measurement requires *memory sensitivity*—that is, the ability to distinguish between those behaviors that occurred and those that did not. A precondition for memory sensitivity is that information must be thoroughly and carefully encoded and retrieved. Yet, both noise and bias may affect follower ratings of leader behavior. *Noise* is a nonsystematic or random error, such as lack of care or fatigue that reduces memory sensitivity for leader behaviors. *Bias* is a nonrandom or systematic response set that reflects a predisposition to respond in a particular manner by either endorsing or not endorsing items.

Reliance on schemas, scripts, and social categories may introduce bias into behavioral ratings of leadership. Raters are prone to endorse behaviors that seem familiar but did not actually happen (e.g., false alarms; Martell & Guzzo, 1991; Shonrick, Dinh, & Lord, 2010), and

such familiarity may stem from the prototypicality of an item within a relevant category. For example, it has long been known that person categorization processes, based on implicit leadership theories, influence ratings of leader behavior. In particular, the same factor structure emerged in ratings of hypothetical leaders as in ratings of familiar leaders (Eden & Leviatan, 1975; Rush, Thomas, & Lord, 1977; Weiss & Adler, 1981), implying that behavioral ratings of leaders may reflect followers' implicit leadership theories rather than actual leader behavior. Importantly, implicit leadership theories can trigger "false alarms" in person perception, whereby observers incorrectly report behaviors that were not observed yet are consistent with the leader prototype (Phillips, 1984; Phillips & Lord, 1982). In this sense, false alarms affect discrimination as well as bias. Martell and Evans (2005) suggest that reliance on prototypical leader behaviors not only produces false memories, but also fosters heightened feelings of familiarity with prototypical leader behaviors, which in turn causes raters to adopt more liberal decision criteria. Similarly, followers' affective reactions to leaders may systematically influence ratings of leader behavior (Bono & Ilies, 2006; Naidoo, Kohari, Lord, & DuBois, 2010) by impacting both memory sensitivity (Allen, Kaut, & Lord, 2008) and bias, as positive affect engenders more liberal thresholds.

Taken together, the evidence suggests that implicit leadership theories and other aspects of raters such as their affective reactions present a significant challenge to ensuring accuracy in follower ratings of leader behavior. Thus, their effects on the data used for theory testing or leadership development represents a critical issue for the leadership field. Accentuating this concern, measures may have little to do with actual leader behavior and are less accurate than we would like them to be, or assume they are (Hunter et al., 2007; Lord, 1985; Yammarino Dubinsky, & Spangler, 1998). For example, Scullen, Mount, and Goff (2000) report that 62% of

the variance in subordinate ratings of leaders on the human dimension, which includes items that pertain to leadership such as "motivates others," "builds relationships," and "listens to others", is associated with idiosyncratic rater effects (e.g., bias). To that end, a comprehensive examination of individual differences that may impact rating accuracy is sorely needed.

In addition, a focus on follower individual differences that impact ratings of leader behavior may inform the future development of leadership studies and constructs such as transformational leadership and leader–member exchange (LMX) that might potentially suffer from endogeneity effects (Antonakis, Bendahan, Jacquart, & Lalive, 2010, 2014) where the effect of x on y cannot be interpreted because of omitted causes. Endogeneity can stem from several factors, including omitted variables, simultaneity, and common-method variance. Concerns about endogeneity are important because if the relationship between x and y is due in part to other reasons, then the correlation simply has no meaning (Antonakis et al., 2010). Thus, endogeneity is a bias-related issue. Individual differences may provide a solution given that they are mostly exogenous to leadership processes and, therefore, can be useful in process models of leadership (Antonakis et al., 2014). As Antonakis, Day, and Schyns observe (2012, p. 644), one promising area for exploration "is to investigate further how follower individual differences affect perceptions of leaders. This research goes back to the idea that the variance in follower leadership ratings is not only a measurement error but also is a reflection of follower individual differences (e.g., Felfe & Heinitz, 2010; Graen, 1975; Hofmann, Morgeson, & Gerras, 2003)." Similarly, Scullen et al. (2000) contend that any causal model seeking to explain ratings of supervisory behavior must account for rater effects, as they are the largest source of rating variance. Clearly, further consideration of the nature of individual and perspective-related effects is needed.

Another important issue concerns the means by which follower ratings of leader behavior are actually produced. We propose that gaining a better understanding of the factors that promote the use of automatic and categorization-based processing is an essential first step to increase accuracy in follower ratings. Similarly, understanding the effects of affect on the rating process is necessary because affective reactions are very fast, often setting the stage for subsequent cognitive processing (Srull & Wyer, 1989). The importance of these issues is clear; when raters rely on automatic processes or extant schemas to simplify processing, many factors associated with those schemas become endogenous to the rating process and, therefore, have the potential to bias estimates of leader behavior effects.

Here, we examine the mechanisms through which follower individual differences, contextual factors, and even the type of measure used may bias ratings of leader behavior; develop a conceptual model that illustrates this process; and offer potential solutions to increase accuracy in follower ratings of leader behavior. Accuracy in ratings of leader behavior is a particularly vexing problem because these ratings represent the end of a highly integrative and ongoing sensemaking process that encompasses leaders, perceivers, and context, and all these factors influence both perception and later retrieval of leader behavior. The person perception and memory literature is used below to illustrate the challenges these sensemaking processes create for accuracy in follower reports of leader behavior.

Accuracy and Bias in Person Perception and Memory

Dual-processing models of person perception (e.g., Brewer, 1988; Fiske & Neuberg, 1990) contend that person perception occurs sequentially, with quick, effortless person

categorization preceding effortful individuation. As detailed by Fiske, Lin, and Neuberg (1999), research has demonstrated that person schemas are central and available to perceivers within milliseconds of encountering another individual, and the automatically guide understanding and encoding of person-related information. Once the target has been categorized, the categorical structure works quickly and efficiently without much effortful thought, eliciting selective perception, interpretation, inference, and memory (Heilman, 1995). Such processing is likely to occur within connectionist systems, and therefore takes place outside of conscious awareness. Perceivers give priority to categorization based on general knowledge structures over individuation or encoding of person-specific information; they will move on to more effortful processing only if targets are judged to be of sufficient motivational relevance and sufficient cognitive resources are available to permit additional processing (Fiske et al., 1999; Gilbert, Pelham, & Krull, 1988). Thus, the information needed to provide accurate ratings or behavioral descriptions of leaders that is independent of other associated effects is typically not available to followers when they are asked to rate leader behaviors.

It should be stressed that this dual information processing system has evolved to benefit the perceiver. It allows for the rapid formation of knowledge-based interpretations as a guide for perceiver behavior, and it enables gradual learning about structures in one's social or physical environment. However, it did not evolve to distinguish behaviors that occurred in a given situation from behaviors that are part of normally expected patterns but did not actually occur (e.g., memory sensitivity). Nor did it evolve to provide ratings that are free from bias. In other words, and of critical importance as it relates to follower ratings of leader behavior, the human memory system was not designed to produce behavioral accuracy in memory, but rather to allow

for meaningful interpretations by observers of social processes and the contexts in which they occur.

As Smith and DeCoster (2000) stress, the patterns extracted from semantic memory are often used as a template to preconsciously process and interpret new information by categorizing and filling in unobserved details through a *pattern completion process*. Semantic memory, therefore, presents significant challenges to accuracy in follower ratings of leader behavior. *The gap-filling function, as mentioned previously, may encourage individuals to report behavior that is consistent with their person schemas and social contexts, yet did not really occur.* In general, such associative or schematic processing is considered the default processing mode, as individuals will engage in more effortful processing only if they are motivated to do so (Fiske & Taylor, 2013). Thus, behavioral accuracy in terms of memory sensitivity may be unusual (Hastie & Park, 1986).

In contrast, biases in describing others that result from the use of classifications and semantic knowledge may be quite common. Indeed, the general problem of accuracy in social perception has long been a problem for social and personality psychologists (Fiske & Taylor, 1984). Moreover, leadership researchers have increasingly recognized its fundamental importance in understanding endogeneity, because the very processes that allow for a quick and meaningful understanding of leadership situations by perceivers may also create many spurious relationships to both ratings of leader behavior and the outcomes those ratings are used to predict.

It is also important to note that for much of everyday activity in organizations, leadership may not be a relevant person-based encoding category to perceivers, as those individuals may be

more interested in competence, warmth, integrity, and trustworthiness as social categories. Thus, even when motivated to accurately perceive others, followers may not encode behaviors that are directly relevant to leadership. When researchers subsequently ask them to provide ratings of leader behavior, respondents may be able to rely only on semantic memory, or general impressions of the target leader, to assess the likelihood of the occurrence of the focal behavior.

Although appropriate classification of social stimuli is an important need that is facilitated by using person schema (Lord, 1985), an even more critical issue is to understand events as they are unfolding and to appropriately gauge one's role in such events. Scripts are knowledge structures that organize events, having both (1) a canonical structure that sequences actions from beginning to end and (2) a hierarchical, goal-related structure that links means to ends. Because perceivers are co-producers of leadership, it is likely that much leadership occurs under the guidance of event-related scripts and, therefore, is encoded using script-related (rather than person-related) structures. In contrast, most measures of leader behavior focus on people (not events), which raises the possibility that much of our research asks for information not typically encoded by raters.

Having described person perception and memory functions in general, we now consider how individual differences might influence these processes. Understanding individual differences that may affect rating accuracy is important because it suggests possible variables that could be included in statistical analyses to control for rater effects on behavioral ratings of leaders. Grounding such individual differences in a carefully developed theory of rater information processing, in addition to rater training or measurement improvements, may offer the best hope for improving the accuracy of follower ratings of leader behavior.

Individual Differences

In this section, we consider individual differences such as personality and attribution styles that have previously been associated with bias or accuracy in person perception and, by extension, with ratings of leader behavior. Individual differences may reflect general tendencies of information processing that ultimately influence ratings. Indeed, rating elevation or bias has been described as a relatively stable rater characteristic (Borman & Hallam, 1991; Kane, Bernardin, Villanova, & Peyrefitte, 1995).

Personality

Big 5. First, *agreeableness* reflects a positive orientation toward others. Agreeable individuals are fundamentally altruistic; they are sympathetic and eager to help others (Costa & McCrae, 1992). Moreover, individuals high in agreeableness are predisposed to view others positively, as agreeableness has been associated with elevated, socially desirable appraisals (Bernardin, Cooke, & Villanova, 2000; Bono et al., 2012). Taken together, the evidence suggests that individuals high in agreeableness may be prone to endorse desirable leader behaviors regardless of their actual occurrence, thereby inserting a positive bias into ratings.

In contrast, the ratings of conscientious individuals are less prone to elevation (Bernardin et al., 2000). *Conscientiousness* is characterized by attention to detail, deliberation, and the tendency to think carefully and cautiously before acting (Costa & McCrae, 1992). Therefore, it seems likely that individuals high in conscientiousness will be less inclined to rely on cognitive short-cuts such as schemas to inform their ratings of leader behavior, but instead will

systematically consider the extent to which leaders display each behavior, thereby producing more accurate ratings.

Accuracy in person perception also requires a willingness to refine and revise initial impressions based on subsequent observations. *Open* individuals have been described as being willing to consider new ideas (Costa & McCrae, 1992) as well as curious about the ideas and attitudes of other people (Welsch, 1975). It seems plausible that open individuals may be both motivated to seek out information about their leaders and willing to revise their impressions if subsequent observations contradict their initial impressions. Therefore, individuals high in openness may provide more accurate ratings of leader behavior. Conversely, individuals low in openness may seize upon the most salient information and dogmatically cling to that information even when they encounter conflicting information (Jost, Glaser, Kruglanski, & Sulloway, 2003).

Not all social information is represented cognitively. Memories of social information also include the emotions associated with a particular person or event. Indeed, Srull and Wyer (1989) argue that emotions impact our initial reactions to individuals as well as guide subsequent information processing. As a result, individual differences in predispositions to experience different emotions may impact attention, interpretation, and subsequent ratings of leader behavior. Both *neuroticism* and *extraversion* have been linked to the frequency with which individuals experience negative and positive affect, respectively (e.g., PANAS; Verduyn & Brans, 2012; Watson & Clark, 1992). On the basis of these strong associations, Tellegen (1985) suggested that neuroticism and extraversion could be renamed negative emotionality and positive emotionality, respectively. We extend the discussion of these traits below by considering how the chronic tendency to experience emotional states may affect rating accuracy.

Positive and Negative Affectivity. Depending on the instructions, positive and negative affect can be measured as either a state or a trait. In a later section, we consider the impact of follower mood states on ratings of leader behavior. Here, we focus on the traits that Tellegen (1985) refers to as negative affectivity (NA) and positive affectivity (PA).

At the trait level, NA is a broad and pervasive disposition to experience negative emotions such as fear, anxiety, hostility, scorn, and disgust that impact one's cognition, self-concept, and worldview (Watson & Clark, 1984). In contrast, PA reflects individual differences in the experience of positive states and engagement with the environment (Watson, Wiese, Vaidya, & Tellegen, 1999). Individuals high in PA report a generalized sense of well-being, competence, and effective interpersonal engagement (Watson, Clark, & Carey, 1988). Although NA and PA might seemingly reflect opposite poles of a single dimension, *when measured at the trait level* they are largely independent, have different correlates, and relate to different classes of variables (Watson, 1988).

Watson et al. (1999) argue that the terms positive affectivity and negative affectivity represent two broader biobehavioral systems of approach and avoidance, respectively. Fowles (1994) has proposed a theory linking NA with the behavioral inhibition system (BIS), which evolved to inhibit behavior that might lead to pain, punishment, or other undesirable consequences. Thus, the primary function of the BIS is to avoid aversive stimuli (Watson et al., 1999). According to Gray (1987), the BIS is a "stop, look, and listen system," in that it redirects attention toward the environment. Such a system promotes an unsettled and vigilant cognitive mode in which the individual scans the environment with uncertainty and apprehension (Gray, 1982, 1985; Tellegen, 1985). Individuals high in NA have an overactive behavioral inhibition system; consequently, they are primed to look for signs of impeding trouble (Gray, 1985).

As a result, individuals high in NA are more likely to notice and attend to normal bodily sensations and aches and pains as well as to interpret such symptoms as painful or pathological (Watson, 1988). Similarly, high NA individuals are better at recognizing and recalling stimuli that have been associated with failure, tend to dwell on and magnify their mistakes (Watson & Clark, 1984), and judge their critics more harshly (Baldwin & Cabianca, 1972). Indeed, several studies have found that such individuals generally interpret ambiguous stimuli in a negative or threatening manner, and are more likely to experience discomfort at all times even in the absence of overt stress (Watson & Clark, 1984). Perhaps not surprisingly, NA is associated with negative views of self and others (Gara, Woolfolk, Cohen, Goldston, Allen, & Novalany, 1993) and lower indices of job, marital, and life satisfaction (Clark, Watson, & Mineka, 1994). Watson (1988) suggests that individuals high in NA may view their life as a series of stressors and hassles regardless of what actually happens to them. A related possibility is that individuals high in NA may be chronically unable to cope effectively with the ongoing events of their lives and, as such, may overreact to them (Depue & Monroe, 1986; Tellegen, 1985). Given these individuals' tendency to interpret ambiguous stimuli more negatively, coupled with their tendency to dwell on and magnify failures, it seems likely that they will be predisposed to perceive feedback from leaders (even neutral feedback) in negative terms. Moreover, NA may serve to increase the availability and salience of negative leader behaviors. Therefore, we anticipate that the ratings of individuals high in NA will reflect a negative bias.

In contrast, trait PA is related to individual differences in the behavioral activation system (BAS), also referred to as the behavioral facilitation system (BFS) (Fowles, 1987; Tellegen, 1985). The basic adaptive feature of the BAS ensures that individuals obtain the resources and cooperation of others that are essential to survival of both the individual and the species (Watson

et al., 1999). Fowles (1987) describes it as "a reward seeking or approach system that responds to positive incentives by activating behavior" (p. 418). Together, the BAS and PA represent biobehavioral systems that encompass incentive motivation, sensitivity to reward signals, interest and alertness, euphoria, excitement, and pleasure seeking (Clark et al., 1994).

PA is associated with enhanced feelings of affiliation, social activity, and satisfaction as well as frequency of pleasant events (Clark & Watson, 1988; Watson, 1988). Based on the idea that material is stored in memory based on its affective tone (e.g., Bower, 1981), Isen (1984) has proposed that individuals high in PA will tend to recall pleasant and rewarding events from their past, which in turn will result in more favorable evaluations of others and expectations of future rewards. By extension, it seems plausible that given their sensitivity to reward signals, individuals high in PA will be predisposed to attend to positive feedback from leaders or even view neutral feedback through a positive lens. Such interpretations may increase the availability of positive leader behaviors. Accordingly, we anticipate that the ratings of individuals high in PA will reflect a positive bias.

Moreover, because NA and PA relate to behavior activation systems, they should have strong relationships to the goal structures that people form in most situations and the type of scripts that they follow. In terms of regulatory focus, PA is likely to be associated with a "promotion" focus, whereas NA is likely to be associated with a "prevention" focus. Regulatory focus is important because it impacts both behavior and encoding of information (Forster, Higgins, & Bianca, 2003) and it has been theorized to have many relations with leader and follower processes (Kark & Van Dijk, 2007). Thus, in addition to PA and NA, individual differences in regulatory focus are likely to affect the way in which information related to leaders and events is encoded. In particular, when processes and people are congruent with an

individual's regulatory focus, they are seen as having more value and as being more correct (Higgins, Idson, Freitas, Spiegel, & Molden, 2003). Perceiver individual differences, therefore, may affect not only the means by which information is encoded, but also the subjective feelings of certainty that leader behaviors consistent with the perceiver's regulatory focus occurred.

Needs and Motives. In addition to personality characteristics, individuals have both basic human needs (e.g., belongingness and attachment) and needs that are specific to organizational settings (e.g., need for leadership). While needs may vary depending on the context, certain reliable, chronic individual differences shape person perception and ratings of leader behavior.

First, *belongingness* can foster automatic processing and reliance on categorization in person perception (Fiske & Taylor, 2013). In leadership settings, the need to be liked and accepted may promote identification with the leader. To that end, self-appraisals may become inextricably linked to leadership appraisals, as individuals are motivated to attribute desirable qualities to leaders and, through reflection, to themselves (Kark & Shamir, 2002). Consequently, the ratings of individuals with high belongingness needs may reflect a positive bias.

Other unmet needs, such as *attachment needs*, may impact perceptions of others and ratings of leader behavior. For example, the attachment system tends to bias cognitive processing in a self-sustaining manner (Mikulincer, Shaver, & Pereg, 2003), such that anxious individuals are biased to view others as potential attachment figures (Mikulincer, Gillath, & Shaver, 2002) and search for even minimal signs of interest and availability. Accordingly, highly anxious individuals view work as an opportunity to satisfy unmet attachment needs (Hazan & Shaver, 1990). In a laboratory study, attachment anxiety predicted ratings of transformational leader

behavior, even though the leader stimulus itself was not transformational (Hansbrough, 2012). As concluded by Hansbrough (2012), it is possible that anxiously attached individuals may be motivated to view leaders as capable of meeting their needs and, therefore, tend to rate them more favorably than do others. Moreover, as long as unmet needs remain salient, anxious followers may be chronically predisposed to view leader behaviors through a positive lens.

In organizational settings, De Vries and Van Gelder (2005) posit that follower needs, such as need for leadership, may impact perceptions of leaders and subsequent ratings of their behavior. *Need for leadership* is the extent to which a follower desires guidance or supervision from his or her leader (De Vries, Roe, & Taillieu, 2002). It should be noted that need for leadership is not a basic need but rather a quasi-need (e.g., Lewin, 1951) that is evoked in a particular setting. Specifically, it becomes relevant within an organizational context, particularly in times of uncertainty or crisis. Followers with a high need for leadership attribute more charisma to their leaders than do other followers (De Vries, Roe, & Taillieu, 1999). Schyns, Kroon, and Moors (2008) also report a positive relationship between need for leadership and ratings of LMX. Moreover, the ratings of followers with a strong need for leadership reflect higher prototypical ratings than those of followers with a weak need for leadership (De Vries, 2000). Based on these findings, we anticipate that followers with a strong need for leadership will be predisposed to report prototypical leader behaviors, such that their ratings will reflect a positive bias.

Attribution Styles. In addition to person perception and categorization, individuals tend to develop accounts for behavior to explain why people do what they do (Heider, 1958; Jones & Davis, 1965; Kelley, 1967). In organizational settings, attributional processes are an important part of the information processing that individuals use to understand the relative importance of

leadership in comparison to other potentially relevant accounts for organizational events (Calder, 1977; Martinko & Gardner, 1987; McElroy & Hunger, 1988; Phillips & Lord, 1981). Like person perception, this explanatory process is far from accurate. Individuals often ignore contextual information about circumstances and instead infer an actor's disposition from his or her behavior (Jones & Davis, 1965). One of the most important insights that emerged from Jones and Davis's (1965) empirical legacy is that people not only wish to infer the dispositions of others, but they are biased to do so (Fiske & Taylor, 2013). Indeed, the fundamental attribution error holds that social perceivers overestimate the impact of personality and minimize the impact of situational factors when providing accounts of targets' behavior. As a consequence, individuals who in reality are constrained by situational factors are held more accountable than they should be (Fiske & Taylor, 1984).

Although the tendency to make attributional errors applies to all individuals, research on attribution styles suggests that some individuals are chronically predisposed to make similar attributions over time (Abramson, Seligman, & Teasdale, 1978; Russell, 1991). Thus, attribution styles reflect a trait-like tendency to display attributional biases such as the actor–observer bias and the self-serving bias (Martinko, Moss, Douglas, & Borkowski, 2007). Causal attributions, of course, are shaped by both individual differences and contextual factors. Contextual factors that may impact attributions and subsequent ratings, such as the follower's physical and psychological distance from the leader, are discussed later in this paper. Here, we focus on how individual differences in attribution styles may inform leadership attributions and ratings of leader behavior.

Some of the earliest work on categorical accuracy and bias in leader ratings adopted an attributional framework. For example, Calder (1977) described leadership as an attribution that

observers use to explain organizational outcomes. Likewise, Pfeffer (1978) described leadership as the outcome of an attribution process in which observers tend to attribute outcomes to the person rather than to the event. By extension, *Romance of Leadership* (ROL) posits that individuals are biased toward using leadership as a mechanism to explain organizational performance in lieu of other possible causes such as situational factors or the followers (Meindl, Ehrlich, & Dukerich, 1985), as leaders are both credited with positive organizational outcomes and blamed for negative ones. According to Meindl (1990), this process is similar to the fundamental attribution error, in that leaders are seen as more powerful, likely causal agents than other factors. Thus ROL is an outcome of an attributional style that reflects a propensity to rely on leadership (e.g., the fundamental attribution error) to explain organizational outcomes.

While the early studies focused on a general tendency to romanticize leadership, Meindl et al. (1985) suggest that the particular attributions will depend at least as much on the characteristics of the observer as on the system itself. Specifically, some individuals, based on their attribution style, may exhibit a dispositional tendency to attribute outcomes to leaders across situations (Meindl, 1990) as evidenced by high scores on the Romance of Leadership Scale (RLS; Meindl & Ehrlich, 1988). For example, Meindl (1990) reports a performance cue effect whereby individuals high in ROL infer strong leadership when presented with evidence of high performance and weak leadership when presented with evidence of low performance. In contrast, the ratings of individuals low in ROL do not vary as a function of performance cues. This result may also reflect the fact that dispositional explanations are required for performance cue effects to occur (Ensari & Murphy, 2003).

Attribution styles that reflect a tendency to explain leader behavior in terms of traits, as indicated by ROL, may also impact the organization and interpretation of information regarding

leaders given that individuals high in ROL may be primed to see heroic traits such as charisma (Meindl, 1990). This assertion has been corroborated by Schyns, Felfe, and Blank's (2007) meta-analysis of 18 studies that reports a positive relationship between Romance of Leadership and ratings of transformational/charismatic leadership. Given that ratings of transformational leader behaviors seem particularly susceptible to bias from attribution style, the interpretation of studies that do not control for attribution style may be problematic. Specifically, it is unclear whether follower reports reflect information about the constructions of the followers or information about the qualities and behaviors of the leader.

Mediating Factors

In this section, we consider individual psychological processes (e.g., mediating factors; Baron & Kenny, 1986) such as stereotype activation and use, perceived similarity, liking, and mood that have previously been offered as explanatory mechanisms for bias or accuracy in person perception and, by extension, ratings of leader behavior.

Stereotype Activation and Use

As discussed previously, person perception begins with affective reactions and reliance on categorization and person schemas. It has long been assumed that people's schematic preconceptions, and their stereotypes, in particular, drive evaluations of and reactions toward others (Allport, 1954; Bartlett, 1932; Tajfel, 1969). Stereotypes have largely been viewed as automatic and unavoidable. For example, Devine (1989) argues that because stereotypes permeate our cultural milieu, they are well known by all individuals even if they do not endorse them. According to this perspective, stereotypes are automatically activated by the mere presence

of a target group member irrespective of any intention on the part of observer (Devine, 1989).

Other studies, however, suggest that stereotype activation is not unconditionally automatic and

may vary as a function of individual attitudes such as overt racism (Blair & Banaji, 1996; Fazio,

Jackson, Dunton, & Williams, 1995; Gilbert & Hixton, 1991; Lepore & Brown, 1997; Locke,

MacLeod, & Walker, 1994). Likewise, egalitarian individuals display no evidence of stereotype

activation following the presentation of a priming categorical cue (Lepore & Brown, 1997;

Locke et al., 1994). Thus, individuals differ in their demonstration of prejudice even at the

categorical level (Wittenbrink, Judd, & Park, 1997). These findings are important because if

stereotypes are not activated (either consciously or unconsciously), then they cannot be used in

subsequent judgments such as leadership ratings. As such, individual differences may influence

the activation and reliance on stereotypes in ratings of leader behavior.

Moreover, even if stereotypes are activated, they need not be applied in judgment and

evaluations. For example, low-prejudice individuals' internal motivation to avoid stereotypic

responses may motivate them to self-regulate their prejudiced responses (Devine, 1989; Devine

& Monteith, 1999). Such observers can quickly switch to more careful and cautious processing

to override automatic stereotypic responses and replace them with more egalitarian responses.

Wegener and Petty (1997) suggest that people vary in whether they believe their judgments are

biased by stereotypes and whether they are motivated to engage in processes to correct for these

biases. In particular, these authors suggest that low-prejudice people who are motivated to

correct for an unfair bias are the individuals who are most likely to manifest such correction

processes. This strategy for overriding stereotypic responses is likely to occur in situations that

call for considered evaluations and judgments of others. Nevertheless, such processes do not

precisely undo the effects of stereotypes, as raters have no reflective access to automatic

processes. Rather, raters who are motivated to be unbiased may overcorrect as well as undercorrect for stereotypes.

Notably, stereotypes are more likely to influence evaluations of others when individuals lack information about actual performance (Terborg & Ilgen, 1975). In particular, observers rely on stereotypes to fill in the gaps of their memory, thereby inserting a systematic bias into ratings of leader behavior. In a later section, we detail how contextual factors such as leader gender and distance may foster reliance on stereotypes in ratings.

Perceived Similarity

The self-concept, which is comprised of self-relevant information, filters perceptions of others (Dunning, 2003; Lord & Brown, 2004; Markus & Wurf, 1987). In particular, individuals who define themselves in terms of a particular characteristic are predisposed to notice that same characteristic in other people (Fiske & Taylor, 1984). In leadership settings, perceived similarity fosters identification with leaders and positive evaluations of leaders (Lord, Brown, & Freiberg, 1999). For example, follower perceptions of similarity predict ratings of leader–member exchange (Engle & Lord, 1997; Liden, Wayne, & Stilwell, 1993; Turban & Jones, 1988). Perceived similarity as described by personality characteristics may also influence ratings of transformational leadership. In particular, ratings of transformational leader behavior have been linked to follower extraversion, agreeableness, openness, and conscientiousness (Bono et al., 2012; Felfe & Schyns, 2010), suggesting that the characteristics associated with transformational leadership may be more salient to individuals who define themselves in similar terms. Bono et al. (2012) conclude, "Rater personality plays a non-trivial role in explaining differences between raters in their reports of leader behavior" (p. 141). We maintain that part of this personality effect

reflects the assimilation of an individual's personality characteristics into his or her self-concept, which in turn affects ratings of leader behavior.

Perceived similarity is traditionally associated with categorization processes and inaccuracy in person perception (Beer & Watson, 2008; Watson, Hubbard, & Wiese, 2000), as self-knowledge acts as a heuristic to automatically fill in the gaps (Human & Biesanz, 2011). Further, this gap-filling process can undercut memory sensitivity for specific types of leader behaviors (Lord, Brown, Harvey, & Hall, 2001; Shondrick et al., 2010). Perceived similarity can be fueled by the motivation to feel connected to others or, as discussed previously, belongingness needs. More generally, assumed similarity facilitates self-esteem maintenance (Marks & Miller, 1987). Therefore, viewing oneself as similar to the leader, coupled with positive ratings of leader behavior, serves as a mechanism to validate one's own personality characteristics.

We anticipate that the ratings of followers who perceive themselves as similar to the leader will reflect a positive bias. Given that perceived similarity enables individuals to sustain positive views of self, it is not surprising that perceptions of similarity promote interpersonal attraction and liking. In short, we like similar others (Berscheid, 1982).

Liking

According to Hunter et al. (2007), liking may bias ratings of leader behavior and, therefore, is a particularly relevant variable to consider in leadership research. Zajonc (1980) argues that affect precedes cognition, rather than vice versa. He posits that affective reactions are basic, inescapable, irrevocable, difficult to verbalize, and implicate the self. In terms of person perception, initial impressions, such as affect or liking, are formed online at the time of the encounter (Bargh & Thein, 1985; Hastie & Park, 1986). Likewise, impression formation models

(e.g., Srull & Wyer, 1989) suggest that person impressions follow a series of stages that begin with an initial overall "general evaluative concept of the person" (e.g., likable or dislikable). This evaluative concept then serves as an interpretative schema that biases subsequent perceptions (Schwarz, 1990; Srull & Wyer, 1989), such that observers selectively attend to information that confirms their initial impressions.

Liking is an important precursor of ratings of leader behavior (Engle & Lord, 1997; Lewter & Lord, 1992; Liden et al., 1993; Wayne & Ferris, 1990). For example, liking assessed as early as the first two weeks of the superior–subordinate relationship predicted follower ratings of LMX six months later (Liden et al., 1993). Because LMX relationships have been shown to form at the pre-entry stage (Liden, Erdogan, & Bauer, 2006), it is reasonable to conclude that liking acts as an antecedent of LMX (Dulebohn, Bommer, Liden, Brouer, & Ferris, 2012). Moreover, Brown and Keeping (2005) report that liking is an important component of the factor structure of the Multifactor Leadership Questionnaire (MLQ) scale.

Taken together, these results suggest that fast-acting affective responses may be an integral component in the evaluation of leaders (Naidoo et al., 2010). Therefore, we anticipate that follower liking will be associated with inflated ratings of leader behavior. As observed by Hunter et al. (2007), such biasing effects of liking are not trivial and may be driving a number of conclusions when the typical leadership study methods are employed (de Groot, Kiker, & Cross, 2000; Fuller, Patterson, Hester, & Stringer, 1996; Lowe, Kroeck, & Sivasubramaniam, 1996). Failure to control for liking may contribute to endogeneity in leadership studies and render subsequent interpretations of data problematic. However, because liking is not exogenous, but rather part of the process that causes leadership, controlling for liking will still not produce consistent estimators for effects of leader behavior (Antonakis et al., 2010).

Mood

Follower mood states may also impact ratings of leader behavior. According to mood-congruent memory, emotion acts as a retrieval cue whereby, memories consistent with an individual's current mood state are more easily recalled. The effects of mood on judgment seem to be caused by the availability of mood-congruent thoughts (Bartlett & Santrock, 1977; Clark & Waddell, 1983; Wright & Mischel, 1982). One of the clearest effects identified in the mood literature is that happy people like just about everything; themselves, their health, other people, the future, and even criminal defendants (for reviews, see Bodenhausen, Mussweiler, Gabriel, & Moreno, 2001; Clore, Schwarz, & Conway, 1994; Crano & Prislin, 2006; Forgas, 1995).

Mood impacts not only the positivity of judgments, but also the manner in which judgments are made (Bodenhausen et al., 2001; Forgas, 1995; Isen, 1993). Individuals in happy moods make decisions quickly (Isen & Means, 1983), group more varied things into the same category (Isen & Daubman, 1984), and make more associations with positive words (Mayer, Mamberg, & Volanth, 1988). In addition, they are often satisfied with quick heuristic judgments (Fiske & Taylor, 2013). By extension, we anticipate that followers in positive mood states may more readily retrieve positive instances of leader behavior, thereby inserting a bias into ratings.

However, mood congruence effects for negative moods are uneven as individuals in these mood states may be motivated to switch from automatic to controlled processing to escape the bad mood (Clark & Isen, 1982). To that end, individuals in negative moods may be more likely to resist automatic associations to negatively toned material and instead rely on controlled processing (Fiske & Taylor, 1984).

Contextual Factors

In this section, we consider how the conditions under which the rating process occurs may influence the accuracy of ratings of leader behavior. In particular, contextual factors such as leader individual differences, distance, national culture, and type of measure used may impact the leader behaviors that are considered salient and are recalled as well as prompt the use of categorization-based processes.

Leader Individual Differences

Physical characteristics of leaders such as facial appearance, height, gender, and race may prompt the use of automatic processing, which may in turn impact ratings of leader behavior.

Facial Appearance. Raters may rely on information gleaned from leaders' faces to inform subsequent ratings. For example, younger-looking individuals are less likely to be endorsed as leaders (Trichas & Schyns, 2012). Raters also attribute characteristics and behaviors to leaders such as warmth, dominance, and trustworthiness based on facial appearance (Little, 2014; Olivola, Eubanks, & Lovelace, 2014). Moreover, attributions of competence from faces of political leaders have been shown to predict the outcomes of elections (Antonakis & Dalgas, 2009). Likewise, Todorov, Mandisodza, Goren, and Hall (2005) report that specific judgments about competence, intelligence, and leadership based on a 1-second exposure to faces predicted the outcome of U.S. congressional elections. Thus, automatic inferences formed during the first encounter with a leader "can have subtle and often subjectively unrecognized effects on subsequent deliberative judgments" (Todorov et al., 2005, p. 1624).

Height. Likewise, height prompts the use of automatic processing. From a socio-biological perspective, height is linked with power and supremacy given that larger individuals

are more likely to win fights (Archer, 1988) and attain social dominance (Ellis, 1994). Height is

considered an important leader characteristic (Werner, 1982) and is associated with leader

emergence in groups (Gawley, Perks, & Curtis, 2009). As Hensley (1993) observes, "the

perception seems to exist that taller individuals are somehow more capable, able or competent"

(p. 40). Accordingly, height impacts ratings of leader behavior. Indeed, Stulp, Buunk, Verhulst,

and Pollet (2013) report that taller U.S. presidents are consistently judged as better leaders with

better communication skills and higher performance than their shorter counterparts. Therefore, it

seems plausible that height may function as a retrieval cue that triggers the accessibility of the

leader prototype, which inflates ratings of behaviors typically associated with the word "leader."

Indeed, "the practice of favoring tall individuals amounts to little more than pure bias" (Judge &

Cable, 2004, p. 438).

Gender. In leadership settings, stereotypes impact ratings of women. First, given that

leadership is stereotypically viewed as a male domain, a bias emerges against female leaders as

they can be seen as a poor fit for such positions by observers (Eagly & Karau, 2002; Lord &

Maher, 1991; Powell, Butterfield, & Parent, 2002). As Heilman and Eagly (2008) observe, the

mismatch between a group stereotype (e.g., women) and a job role (e.g., leader) fuels negative

performance expectations, which in turn produces biased evaluations. These performance

expectations affect both how women are regarded and how they are evaluated. Specifically,

women are rated as less competent, less influential, and less likely to have played a leadership

role than their male counterparts (Heilman & Haynes, 2005). Likewise, raters are less likely to

recognize agentic behavior or endorse behaviors typically associated with leaders when the target

is female rather than male (Martell, 1996; Scott & Brown, 2006). Thus, when all else is equal,

female leaders are generally perceived as less effective than male leaders (Johnson, Murphy, Zewdie, & Richard, 2008).

Race. In addition to gender, race impacts ratings of leader behavior. For example, raters perceive non-whites as less effective leaders and less likely to succeed than whites. Moreover, social perceivers give non-whites less credit for their success and hold lower expectations for their future success (Rosette, Leonardelli, & Phillips, 2008). Similarly, Asian Americans are significantly less likely to be viewed as leaders than Caucasian Americans (Sy et al., 2010). Thus a mismatch between racial stereotypes and the job role (e.g., leader) may promote a bias against non-white leaders, whereby they are given consistently lower ratings.

Distance

The notion that close relationships between leaders and followers foster better organizational outcomes is integral to many leadership theories, such as LMX, yet leadership research has only recently begun to explicitly examine distance (Popper, 2013). Leader distance includes physical distance, interaction frequency, and social distance which affect both social intimacy and contact between leaders and followers (Antonakis & Atwater, 2002). Popper (2013) extends this definition to include a subjective aspect, referred to as psychological distance, which is impacted by emotions, construal of social information, and perceptual biases or attributions.

In terms of person perception, people construe distant entities more abstractly than close ones (Liberman, Troupe, & Stephan, 2007). Thus individuals rely on simplified prototypes to inform their perceptions of distant others (Fiske & Taylor, 2013; Smith, 1998). Likewise, Popper (2013) contends that followers rely on trait-based categorization processes to inform their perceptions of distant leaders. Moreover, when followers perceive leaders as psychologically

close, they are more likely to attend to specific leader behaviors and, in turn, demonstrate increased memory sensitivity for those behaviors. This process is consistent with Shamir's (1995) argument that the characteristics attributed to physically distant leaders are more schema based than those attributed to close leaders.

Greater distance may also influence attributions for leader success or failure. Specifically, the tendency to romanticize leaders may increase as the perceived social, interactional, physical, and geographical distance between leaders and followers increases (Antonakis & Atwater, 2002), whereas followers who have closer relationships with their leaders are likely to have greater knowledge of the situational factors that may impact leader performance. Taken together, the evidence suggests that the ratings of followers who perceive leaders as close will tend to reflect greater behavioral level accuracy. In contrast, the ratings of followers who are physically distant from leaders are more likely to be shaped by schemas and categorization processes as there is little information available to contradict their initial impressions. Moreover, physical distance limits opportunities for followers to actually witness important leadership activities, thereby reducing the accuracy of ratings of leader behavior (Hunter et al., 2007).

Although physical distance may certainly impact rating accuracy, it is unclear whether it positively or negatively biases ratings of leader behavior. For distance to positively bias ratings, followers must first categorize their supervisor as a "leader." It seems likely that individual differences and mediating factors, such as follower needs and stereotype activation and use, may determine whether a particular follower is likely to view his or her supervisor as a leader. Once a target has been categorized as a leader, followers may rely on implicit leadership theories to fill in the gaps in their knowledge which serves to inflate ratings. However, other followers may be prejudiced against viewing a particular supervisor as a leader and, in turn, categorize that

supervisor as a "non-leader." This type of categorization may introduce a negative bias into ratings of leader behavior.

National Culture

      Culturally Contingent Implicit Leadership Theories. The cultural background of the perceiver may serve as a lens that filters the interpretation of the social environment (Den Hartog, House, Hanges, & Ruiz-Quintanilla, 1999). Culture may impact ratings of leader behavior in many respects including the types of leader behaviors that are considered socially desirable, the type of information that is attended to and later recalled, and causal attributions for behavior.

      The Global Leadership and Organizational Behavior Effectiveness (GLOBE) research program found that while leader attributes associated with charismatic/transformational leadership are universally endorsed, the endorsement of other leader attributes—such as risk taking, ambition, compassion, enthusiasm, sensitivity, self-sacrifice, and willfulness—is culturally contingent (Den Hartog et al., 1999). Hanges, Lord, and Dickson (2000) propose that some leader behaviors may activate well-established cultural meaning patterns that provide a frame which triggers specific traits and behaviors associated with leadership. Once activated, this frame may inflate ratings of leader behavior, with individuals being inclined to endorse behaviors that fit the culturally endorsed prototype regardless of whether or not they actually occurred.

      Culture and Information Processing. Culture can also trigger self-concepts, such as the independent self and the interdependent self, that impact information processing. The independent self emphasizes a person's uniqueness and internal beliefs, whereas the

interdependent self emphasizes the connection between the person and social relationships (Markus & Kitayama, 1991). This distinction is important because it affects the type of information that is deemed salient and recalled. In particular, individuals who have interdependent selves in working memory are more likely to remember information about the behavior and actions of others (Hanges et al., 2000). Therefore, we anticipate that the ratings of leader behavior by individuals who have interdependent selves in working memory will tend to be more accurate than the ratings of individuals who have independent selves in working memory.

Culture and Attribution Style. Finally, culture impacts attribution style. For example, individuals from collectivistic cultures engage in more holistic thinking and give greater weight to explanations of behavior rooted in situational or contextual factors than individuals from individualistic cultures (Ji, Peng, & Nisbett, 2000; Norenzayan, Choi, & Nisbett, 1999). Consequently, the fundamental attribution error is more widespread in individualistic cultures, where emphasis is placed on the person rather than on the individual in relation to the group (Morris & Peng, 1994). By extension, we anticipate that individuals from collectivistic cultures will be less susceptible to the Romance of Leadership.

Research Methods and Bias

Here, we consider how the research methods used may inadvertently introduce bias into ratings of leader behavior. In particular, accuracy of behavioral ratings is impacted by the type of memory triggered by the questionnaire used to create those ratings. Moreover, data gathering methods such as participant observation, interviews, and experience-based sampling procedures may introduce additional biases into ratings.

Type of Questionnaire. Most types of measures used to assess leader behavior rely on questionnaires distributed to peers or followers (Hunter et al., 2007). Typically raters are asked to provide retrospective frequency judgments for each item that reflects typical leader behavior. It is assumed that such judgments correspond to actual behaviors, although, as we have noted, they may also reflect rater categorization processes and may be affected by many types of individual differences. Such measures likely reflect accurate categorization of leaders in terms of an underlying categorical structure but are unlikely to reflect accurate accounts of how often specific types of leader behavior occurred (Lord, 1985). Moreover, the type of measure used may exacerbate the tendency to rely on schematic-based processing as recognition-based approaches trigger the use of prototypes and fuzzy sets whereby the target is compared to the typical leader (Shondrick et al., 2010). Because they rely on patterns of information aggregated across many events, such ratings encourage individuals to draw from categorization-based processes (Foti & Hauenstein, 2007; Smith & Foti, 1998).

Further, encoded behavior is often stored in a compressed, abstracted form (McClelland, McNaughton, & O'Reilly, 1995), rather than in terms of rich details. Details may be added at retrieval, subject to contextual information, the needs of raters, or their expectations, but these details may not correspond to the original behaviors. This process may activate recall of elements of patterns that never actually occurred, yet are consistent with the pattern (e.g., categorization), thereby inflating estimates of the frequencies of observed leader behavior. Moreover, such measures may be affected by the constructs (e.g., group performance) that measures are used to predict, as knowledge of positive group performance both inflates leader ratings (Binning & Lord, 1980; Rush, Phillips, & Lord, 1981; Rush et al., 1977) and impacts ratings of many types of group processes (Staw, 1975).

Ultimately, rated behavior has many variance components that are endogenous to the rating system. As we detail later, to more accurately measure actual leader behavior, ratings may need to focus on explicit events that occur in a specific context (Shondrick et al., 2010) and prompt raters to use event schemas rather than person schemas (Foti & Lord, 1987).

Participant Observation. In much leadership research, the role of rater information processing is ignored as it relates to the measurement of leader behavior. Nevertheless, such factors have been clearly recognized when researchers are active participants in the process being measured (LeCompte & Goetz, 1982). For example, it is likely that the personalities, needs, and experiences of participant observers affect the schemas they use to interpret a situation and consequently have a large impact on what is later recorded or described. All "first-hand" accounts may be vulnerable to individual differences effects as described heretofore. Their organization into cases, rather than measures of leader behavior, does not eliminate the potential for such effects, although case-based descriptions may tend to align with script-based structures.

Interviews. Although most reports of leader behavior require no direct input from the leader being rated, some types of measures, such as interviews, rely on direct input from the leader. Although an interview format allows for substantial flexibility in addressing process and content issues related to leadership, the responses obtained using this format may be affected by the researcher's own information processing. In particular, the questions asked by researchers may unwittingly prompt leaders to respond in a way that confirms the researchers' expectations (e.g., demand characteristics; Orne, 2009). Thus, more bias is likely introduced as processes become less structured.
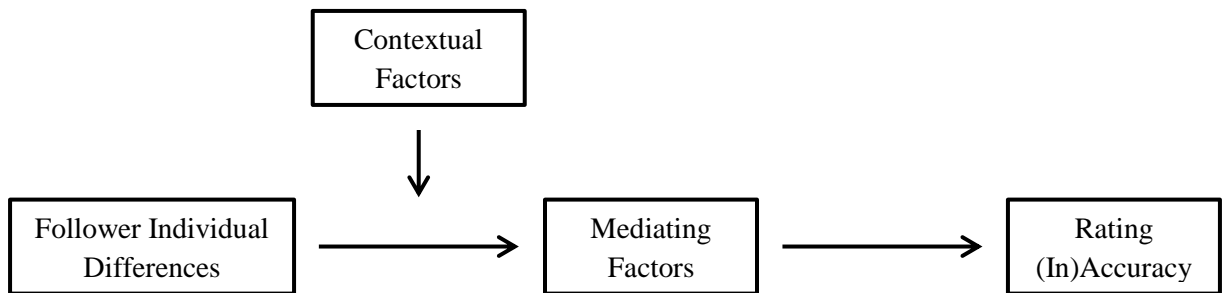
Experience-Based Sampling Procedures. Finally, we should note that one advantage of a typical questionnaire is that, because it is memory based, it cannot directly affect the actual process or leader being described. This contrasts with experience-based sampling procedures, which often measure a specific process multiple times in a day or week. Such measures may not only have carry-over effects that influence subsequent measurement, but also could affect the actual processes that are being measured. This is particularly likely when measures activate important needs or values of the individuals being queried.

Conceptual Model

As depicted in the Figure 1, accuracy in follower ratings of leader behavior is impacted by several factors that include (1) what followers bring to the rating process in the form of individual differences, (2) psychological processes that may be activated within followers at the time of the rating process, and (3) the context in which the ratings take place. First, follower individual differences, such as personality and attribution styles, reflect general tendencies and patterns of perceiving and processing information that with repeated use become more efficient and automatic. Thus follower individual differences impact both the availability and encoding of information. Second, psychological processes within followers (e.g., mediating variables), such as stereotype activation and use, perceived similarity, liking, and positive moods, may prompt reliance on automatic, categorization-based processing. Third, contextual factors, such as leader individual differences, distance, national culture, and the type of measure used, may activate the information (e.g., leader behaviors) that is deemed salient and recalled. Moreover, such

contextual factors may prompt followers to rely on person categorization to inform their

responses, which reduces accuracy at the behavioral level.

Figure 1: Conceptual Model

```
                    ┌──────────────┐
                    │  Contextual  │
                    │   Factors    │
                    └──────┬───────┘
                           │
                           ▼
┌──────────────┐    ┌──────────────┐    ┌──────────────┐
│   Follower   │    │              │    │              │
│  Individual  │───▶│  Mediating   │───▶│   Rating     │
│ Differences  │    │   Factors    │    │ (In)Accuracy │
└──────────────┘    └──────────────┘    └──────────────┘
```

It should be noted that bias may be introduced at several stages in the rating process.

Therefore, follower ratings of leader behavior represent the end of a highly integrative

sensemaking process. As such, the end product may bear little resemblance to actual leader

behaviors.

Discussion

Because leadership involves behaviors, traits, and characteristics as interpreted by

observers (Calder, 1977; Lord & Maher, 1991), the study of leadership requires consideration of

followers' person perception and information processing (Brown & Lord, 2001). Likewise,

Mount and Scullen (2001) observe that ratings of leader behavior largely reflect the idiosyncratic

tendencies of the rater; thus, it is imperative to better understand how and why raters form their

unique interpretations. To that we add that it is important to examine how individual differences

and the contextual factors in which ratings take place might encourage followers to rely on automatic categorization-based processes that in turn reduce accuracy in ratings. Accordingly, our work provides a comprehensive analysis of what followers contribute to the measurement process by identifying how reliable individual differences, internal psychological processes, contextual factors, and type of measure used may contribute to bias in ratings of leader behavior. In this sense, our work answers Brown and Lord's (2001) call for more research that focuses on understanding the subordinate information processing that underlies responses to leadership questionnaires. We close by offering potential solutions for improving rating accuracy.

Solutions for Improving Rating Accuracy

1. *Given all the potential biases in follower ratings of leader behavior, are you suggesting that followers not be used as a source of information?*

No, rather our intention is to call for a more complete understanding of how follower individual differences, internal psychological processes, and contextual factors influence rating accuracy. A better understanding of the limitations of follower ratings of leader behavior allows leadership researchers to make more strategic decisions about theory and measurement as well as to be more precise in the interpretation of their studies. On the one hand, if we are interested in followers' *perceptions of leader behavior*, the issue of bias is irrelevant, because each follower report is interpreted as one individual's perspective or unique experience. On the other hand, if our objective is to identify *leader behaviors* that are associated with organizational effectiveness, as is often the case with many leadership theories and leadership development interventions, then attention to the factors that may bias ratings of leader behavior is essential. Moreover, in terms of testing leadership theory, careful consideration should be given to the individual differences that

are known to impact ratings to help minimize endogeneity (e.g., Antonakis et al., 2010).
Otherwise, the interpretation of those ratings becomes challenging. Despite such concerns, follower ratings of leader behavior are routinely interpreted as identifying effective leader behaviors instead of idiosyncratic perceptions about leaders, which prevents us from developing an accurate understanding of a leader's effects on performance.

2. *What specific recommendations do you offer for improving the accuracy of follower ratings of leader behavior?*

Researchers may wish to consider several options to increase the accuracy of follower ratings of leader behavior: (1) solutions that focus on the followers, such as rater training and selection; (2) solutions that focus on data collection and measurement considerations, such as collecting multiple reports over time and increasing the use of measures that tap into episodic memory and/or creating conditions conducive to the recall of episodic memory; and (3) solutions that focus on data analysis such as controlling for individual differences that are known to impact rating accuracy.

Solutions That Focus on Followers. Rater training might improve accuracy in follower ratings of leader behavior. For example, Martell and Evans (2005) trained raters to report only behaviors that evoked detailed memories and to avoid behaviors based on feelings of familiarity. The results of this source monitoring training substantially reduced bias in behavioral measurement, but it did not increase memory sensitivity. Furthermore, the effects of training may depend on the type of accuracy required by the theory (e.g., categorization or behavioral-level accuracy). Sulsky and Day (1994) note that frame of reference training can increase the tendency of raters to use online encoding that stores a general evaluation but not the underlying details,

thereby increasing accuracy in terms of the frame of reference structure, but decreasing accuracy in terms of rating specific behaviors. Therefore, frame of reference training increases the availability of prototypical leader behaviors. To be most effective, training needs to center on clear examples that illustrate the leader behavior in question, as prototypes have a clearly defined center but fuzzy boundaries. "In short, frame of reference training may promote effective on-line or inference-memory based processing (Hastie & Pennington, 1989), but an unintended consequence could be the general failure on the part of raters to encode and store behavioral information in a manner that facilitates later recall" (Sulsky & Day, 1994, p. 542).

The use of stereotypes during encoding can be reduced by providing raters with individuating information (Heilman & Haynes, 2005; Landy, 2008), explicit instructions to avoid stereotypic responses (Macrae, Milne, & Bodenhausen, 1994), or instructions that stress the importance of accuracy (Brewer, 1988; Fiske & Neuberg, 1990) and therefore promote more careful, controlled processing. However, Heilman and Eagly (2008) caution that although individuating information can deter stereotypes, there are many circumstances in which it does not have that effect as the expectations that stereotypes produce can be tenacious even in the face of disconfirming evidence. Perhaps the most effective means of reducing individual-level stereotypes is to educate individuals about how unconscious stereotyping can occur (Operario & Fiske, 2001), as a better understanding of these processes is the first step to controlling them. Finally, given that individuals are more likely to use stereotypes and categorization-based processes as a gap-filling function when they lack information about actual performance (e.g., Antonakis & Atwater, 2002; Shamir, 1995; Terborg & Ilgen, 1975), researchers may wish to carefully consider whether to include the ratings of followers who may have had little opportunity to observe the leader behaviors in question (Hunter et al., 2007). Conversely, followers who have had greater contact

with leaders may be less inclined to rely on stereotypes and categorization processes. Thus consideration of distance may allow for more strategic choices about sample selection and ultimately increase rating accuracy.

Solutions That Focus on Data Collection and Measurement. Given that leadership is a dynamic process, the behavioral-level accuracy of ratings may change over time. For example, dual processing models in person perception suggest that although individuals initially rely on automatic categorization processes (Brewer, 1988; Fiske & Neuberg, 1990), depending on their motivation, they may later switch to more effortful processing. Moreover, as followers are co-producers of the leadership process (Shamir, 2007), it is possible that what began as rating error might later result in confirming behavior from some leaders. For example, followers with a high need for leadership may not only project their needs onto leaders (Schyns et al., 2008), but over time may also elicit behavior from responsive leaders that satisfies their needs. In short, while needs and motives may bias ratings of leader behavior, they may also change leadership processes and ultimately even the follower needs that prompted the development of a particular dyadic relationship. Because accuracy in ratings may vary over time as individuals either gain more information about leaders or elicit behavior from leaders that confirms their initial impressions, multiple reports may provide greater information about individual-specific patterns of behavior.

The type of measure used can also promote increased accuracy in ratings of leader behavior at the point of retrieval. In cases where leadership is assumed to reflect a group-level process, researchers may wish to develop measures that focus on widely experienced critical incidents or to create conditions that help elicit episodic memory to foster accuracy at the behavioral level. The urgency around this issue is illustrated by concerns about the extent to

which transformational leader behaviors are "in the eye of the beholder (Bono et al., 2012; Brown & Keeping, 2005; Jung, Yammarino, & Lee, 2009; Yammarino et al., 1998). Thus generalizing from one rater to another is problematic because raters do not agree much with each other (Mount & Scullen, 2001). The use of measures that tap into alternative types of structures such as scripts may yield greater behavioral accuracy (Foti & Lord, 1987), inter-rater agreement, and perhaps more insightful leadership theories (Morgeson, 2005). Interestingly, encoding in terms of scripts rather than person-schemas orients one toward situations and reduces the overemphasis on person-based explanations. For example, Morgeson (2005), using a previous specific problem or event, found substantial group-level agreement in ratings of leader effectiveness and satisfaction with the way the leader responded to the event. To that end, researchers could increase behavioral-level accuracy in ratings of leader behavior by using items that ask about specific events (i.e., critical incidents) in lieu of items that rely on general impressions or semantic memory. It may also be worth considering whether event taxonomies (Hoffman & Lord, 2013) could form the basis for event-related measures of leader behavior.

As discussed previously, ratings based on episodic memory may be more likely to accurately reflect leader behavior (Martell & Evans, 2005) and less subject to bias than ratings based on semantic memory (Shondrick et al., 2010). Moreover, Martell and Evans (2005) have shown that training raters to emphasize episodic memory reduces the biases associated with implicit leadership theories. Such research suggests that it may be worthwhile to consider whether an item is primarily based on semantic memory or episodic memory when constructing measures of leader behavior. Tulving (1985) suggests that individuals have metacognitions about memory and retrieval processes that enable them to distinguish between "know" memory judgments (a general feeling or impression of a person—a semantic memory) and "remember"

judgments (a vivid recollection of a specific event—an episodic memory). Items that focus on a rater's role in the leadership process, his or her goals during a specific event, or felt emotions might be especially likely to reflect episodic encoding and retrieval.

Alternatively, researchers could help elicit episodic memory by creating conditions that tap into the emotional response surrounding a prior event. Use of episodic memory is facilitated by events that have high emotional content (Allen et al., 2008). For example, Naidoo et al. (2010) used a visualization procedure that triggered affect prior to ratings which increased rating accuracy. Thus triggering the emotion associated with a particular event may facilitate memory for specific details about events.

However, episodic memory is not a panacea. Attention is highly selective; many details simply go unnoticed although they may shape reactions in an implicit or automatic manner. Thus, even episodic memory which is rich in contextual, self-relevant details is undergirded by long-term memory structures that allow individuals to notice and understand various patterns such as those that are self-relevant or self-reflective. Consequently, even when episodic memories are easily retrieved, they may be oriented toward self-relevant details, such as one's goals, needs, emotions, and identities, rather than the constructs that scientists want to assess via the memory processes of raters. Therefore, it is still important to consider the impact of individual differences on the rating process.

A related issue centers on the definition of the construct of leadership. One problem with the science of leadership is that we often take constructs developed at one level of analysis (e.g., individual relations with leaders) and apply them at other levels of analysis (e.g., the group or organizational level) without recognizing that different constructs may be needed as the level of

analysis changes (Klein, Dansereau, & Hall, 1994). This is not a question of agreement among raters when we aggregate, but rather a theoretical question of whether the same leadership constructs are meaningful at different levels of analysis. This issue becomes less problematic if leadership is defined as a function of leader traits. In such a case, leader behavior would be expected to be consistent across time and across followers. In contrast, if leadership is defined as a system that reflects leaders, followers, and the context, then shifting levels of analysis may pose significant conceptual challenges. This latter approach reflects a more science-based view of leadership, whereas a general trait view aligns better with "common-sense views of leadership". As Calder (1977) noted, when researchers use factor analysis as a basis of construct development, they are likely to tap into "common-sense views of leadership" based on the implicit theories of raters. Implicit theories not only affect the accuracy of ratings, but may also have secondary effects by influencing which concepts are investigated, how those constructs are generalized, and thus, may miss key ideas. Moreover, this suggests that a reductionist approach to measurement development strips leader behaviors of their rich, contextual details. Therefore, raters must rely on generalized impressions regarding leader behavior and the frequency with which it occurred. As an alternative, science-based approaches to generate leadership constructs may help alleviate some of the rater-related issues we have discussed.

Both behavioral questionnaires and questionnaires assessing leader–follower relations ask for explicit assessments of a domain represented by a specific item. In contrast, implicit measures use different response formats, such as word fragment completion and reaction times, and are geared more toward gauging the accessibility of that domain, or its association with various constructs (Uhlmann, Leavitt, Menges, Koopman, Howe, & Johnson, 2012). Because such measures pertain to the accessibility rather than the content of the schema, they may be less

affected by social desirability, but may still reflect categorization processes (Dinh & Lord, 2012). Implicit measures can also assess the tendency of individuals to process patterns of attributes when they categorize an individual (Dinh & Lord, 2012), a phenomenon that applies to explicit measures as well (Foti & Hauenstein, 2007).

Implicit measures may have a special value in addressing endogeneity through the use of ratings of leader behavior to predict organizational outcomes. Rather than direct evaluation of an item's content, implicit measures reflect the accessibility of the underlying construct or its association with other aspects of the rater's semantic memory. Thus, they may have value as very general "measured common causes" estimates. Such a control could conceivably eliminate many endogeneity effects that operate through the effects of individual differences on semantic networks (see Antonakis et al., 2010, Figure 4C). However, implicit measures would not be effective control variables for conscious processes associated with individual differences.

Solutions That Focus on Data Analysis. As this review has illustrated, many individual differences may impact the accuracy of ratings of leader behavior. Therefore, researchers may wish to control for some of these variables when testing their theories. However, it should be noted that controlling for individual differences does not entirely eliminate rater effects and thereby create consistent estimates of the effects for rated leader behavior. As the breadth of this review demonstrates, it may be difficult to include all of the individual difference factors as controls, raising the possibility that the coefficients for rated leader behavior will still not be consistent because of the omitted variable problem (Antonakis et al., 2010). More troubling is the notion that followers are part of the system that produces leader behavior, and the very individual differences we want to use as a means to create unbiased estimates of the effects of leader behavior may also contribute to the process we wish to measure.

3. *Does this issue apply only to behavioral measures, or can it be extended to concepts such as LMX?*

Measures that focus on aspects such as one's relationship with a leader (Liden & Maslyn, 1998) or the degree of trust one has in a leader (Mayer, Davis, & Schoorman, 1995) often have similar formats to behavioral questionnaires, but they ask about the raters' current assessments of their relationship with their leader, rather than the past behavior of the leader. Such measures are particularly likely to reflect individual differences in raters and, therefore, influence the extent to which they are correlated with the particular relationship being measured. Thus, relationship assessments suffer from the same problems as predictors of outcome variables—namely, they are likely to yield biased estimates due to endogeneity. This problem does not arise when relationship variables are used as dependent variables in predictive equations.

4. *Are you suggesting that researchers control for all these individual differences? If so, would there be any variance left? Moreover, if individual differences are normally distributed, wouldn't the variance just cancel out?*

No, as indicated above controlling for all possible individual differences may be difficult. However, we do recommend that researchers err on the side of caution and include more, rather than fewer, control variables (Cameron & Trivedi, 2005). The selection of which variables to include should be guided by "theory, theory, and more theory" (Antonakis & Dietz, 2011, p. 218). It is unclear whether individual differences would cancel out; in order to do so, unique rater variance would have to be random and/or the effects of rater individual differences would have to be randomly distributed across leaders (Bono et al., 2012). Yet Bono et al. (2012), in a study on the impact of rater personality on ratings of transformational leader behavior, report that

unique rater variance is not random and can be attributed to rater personality. Moreover, rater personality is not randomly distributed across leaders, which suggests that "aggregating results to the leader level to predict outcomes, on the assumption that the results would generalize to other work-groups may be unwise" (Bono et al., 2012, p. 139).

5. *What about other rater individual differences, such as gender?*

Demographic variables such as age, race, and gender may impact information processing and ratings of leader behavior. In particular, while some studies suggest that women tend to rate others higher than do men (e.g., Hamner, Kim, Baird, & Bigoness, 1974; London & Poplawski, 1976), other studies report no evidence of gender differences (Lee & Alvares, 1977). Likewise, in an organizational setting, Shore and Thornton (1986) found no evidence of gender differences and concluded that "previous laboratory findings about gender differences in performance ratings may have limited generalizability to the workplace" (p. 115). Given that our objective here is to identify reliable individual differences that impact ratings of leader behavior, a discussion of follower demographic variables is beyond the scope of this paper. Future research may wish to explore how the rater's internalization of gender stereotypes (e.g., women are nurturing, warm, and supportive; O'Leary, 1974) as well as contextual factors may impact perceptions of leaders and subsequent rating accuracy.

6. *If leadership is ultimately in the eye of beholder, does accuracy in ratings of leader behavior matter?*

It may, depending on the objective of the study. If the purpose of the study is to predict individual-level outcomes such as job satisfaction or satisfaction with the leader, then tapping into individual-level perceptions may provide meaningful data as long as the reports are

interpreted to reflect one individual's unique perspective, which may or may not be shared by others. In contrast, it should be stressed that scientific theories of how leaders affect performance of their unit need behavioral-level accuracy to disentangle the effects of generalized perceptions from the effects of leaders on unit performance. A chronic problem in the leadership area is that we have a limited understanding of precisely how leaders affect performance (Kaiser, Hogan, & Craig, 2008), in part because measures are not specific enough to separate behavioral effects from a general classification of an individual as a leader. Better leadership theory may require event-level measurement that is accurate at the behavioral level (Hoffman & Lord, 2013).

Whether accuracy matters also depends on how the ratings will be used. If ratings of leader behavior are used for leadership development purposes, then accuracy in ratings is important because the training based on those ratings may not transfer to other situations or followers.

Conclusion

Accuracy in follower ratings of leader behavior presents a significant challenge to leadership research because it is affected by many factors, including follower individual differences, internal psychological processes, contextual factors, and type of measure used. Some of the most popular and contemporary approaches to leadership have not explicitly considered the role of person perception processes (Brown & Lord, 2001), yet the scientific study of leadership requires a greater sensitivity to followers' information processing beyond the traditional focus on reported leader characteristics and actions. To that end, an enhanced understanding of the factors that impact follower information processing, retrieval of memories

related to leader behavior, and subsequent ratings of leader behavior may better inform

leadership theories, measurement development, and interpretation of results.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87,* 49–74.

Allen, P., Kaut, K., & Lord, R.G., (2008). Emotion and episodic memory. In E. Dere, A. Easton, L. Nadel, & J. P. Huston (Eds.), *Handbook of behavioral neuroscience: Episodic memory research* (Vol. 18, pp. 115–132). New York: Elsevier Science.

Allport, F. H. (1954). *The nature of prejudice.* Reading, MA: Addison-Wesley.

Antonakis, J., & Atwater, L. (2002). Leader distance: A review and proposed theory. *Leadership Quarterly, 13,* 673–704.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly, 21,* 1086–1120.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. In D. V. Day (Ed.), *The Oxford handbook of leadership and organizations* (pp. 93-117). New York, NY: Oxford University Press.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science, 323*(5918), 1183.

Antonakis, J., Day, D. V., & Schyns, B. (2012). Leadership and individual differences: At the cusp of a renaissance. *Leadership Quarterly, 23,* 643–650.

Antonakis, J., & Dietz, D. (2011). More on testing for validity instead of looking for it. *Personality and Individual Differences, 50,* 418–421.

Archer, J. (1988). *The biology of aggression*. Cambridge, UK: Cambridge University Press.

Baldwin, B. A., & Cabianca, W. A. (1972). Defensive strategies of repressors and sensitizers in counseling. *Journal of Counseling Psychology, 19,* 16–20.

Bargh, J. A., & Thein, R. D. (1985). Individual construct accessibility, person memory, and the recall–judgment link: The case of information overload. *Journal of Personality and Social Psychology*, *49*(5), 1129–1146.

Baron, R. A., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Bartlett, J. C., & Santrock, J. W. (1977). Affect-dependent episodic memory in young children. *Child Development, 50,* 513–518.

Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment*, *90*(3), 250–260.

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85,* 232–244.

Berscheid, E. (1982). Attraction and emotion in interpersonal relationships. In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition: The 17th Annual Carnegie Symposium on Cognition* (pp. 37–54). Hillsdale, NJ: Erlbaum.

Binning, J. F., & Lord, R. G. (1980). Boundary conditions for performance cue effects on group process ratings: Familiarity versus type of feedback. *Organizational Behavior and Human Performance, 26,* 115–130.

Blair, I. V., & Banaji, M. R. (1996) Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70,* 1142–1163.

Bodenhausen, G. V., Mussweiler, T., Gabriel, S., & Moreno, K. N. (2001). Affective influences on stereotyping and intergroup relations. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 319–342). Mahwah, NJ: Erlbaum.

Bono, J. E., Hooper, A. C., & Yoon, D. J. (2012). Impact of rater personality on transformational and transactional leadership ratings. *Leadership Quarterly*, *23*(1), 132–145.

Bono, J. E., & Ilies, R. (2006). Charisma, positive emotions and mood contagion. *Leadership Quarterly*, *17*, 317–334.

Borman, W. C., & Hallam, G. L. (1991). Observation accuracy for assessors of work-sample performance: Consistency across task and individual-differences correlates. *Journal of Applied Psychology, 76,* 11–18.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148.

Brewer, M. B. (1988). A dual-process model of impression formation. In R. S. Wyer & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, D. J., & Keeping. L. M. (2005). Elaborating the construct of transformational leadership: The role of affect. *Leadership Quarterly*, *16*(2), 245–273.

Brown, D. J., & Lord, R. G. (2001). Leadership and perceiver cognitions: Moving beyond first order constructs. In M. London (Ed.), *How people evaluate others in organizations* (pp. 181-202). Mahwah, NJ: Lawrence Erlbaum Associates.

Calder, B. J. (1977). An attribution theory of leadership. In B. Staw & G. Salancik (Eds.), *New directions in organizational behavior* (pp. 179–204). Chicago, IL: St. Clair Press.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications.* New York, NY: Cambridge University Press.

Clark, M. S., & Isen, A. M. (1982). Toward understanding the relationship between feeling states and social behavior. In A. Hastorf & A. Isen (Eds.), *Cognitive social psychology* (pp. 73–108). New York, NY: Elsevier North-Holland.

Clark, M. S., & Waddell, B. A. (1983). Effects of moods on thoughts about helping, attraction and information acquisition. *Social Psychology Quarterly, 46,* 31–35.

Clark, L. A., & Watson, D. (1988). Mood and the mundane: Relations between daily life events and self-reported mood. *Journal of Personality and Social Psychology, 54,* 292–306.

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103,* 103–116.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition: Vol. 1 Basic processes* (2nd ed., pp. 323–417). Hillsdale, NJ: Erlbaum.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.

Crano, W. D., & Prislin, R. (2006). Attitudes and persuasion. *Annual Review of Psychology, 57,* 345–374.

de Groot, T., Kiker, S. D., & Cross, T. C. (2000). A meta-analysis to review organizational outcomes related to charismatic leadership. *Canadian Journal of Administrative Sciences, 17*, 356–371.

Den Hartog, D. N., House, R. J., Hanges, P. J., & Ruiz-Quintanilla, S. A. (1999). Culture specific and cross-culturally generalizable implicit leadership theories: Are attributes of charismatic/transformational leadership universally endorsed*? Leadership Quarterly, 10,* 219–256.

Depue, R. A., & Monroe, S. M. (1986). Conceptualization and measurement of human disorder in life stress research: The problem of chronic disturbance. *Psychological Bulletin, 99,* 36–51.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 339–360). New York, NY: Guilford Press.

De Vries, R. E. (2000). When leaders have character: Need for leadership, performance, and the attribution of leadership. *Journal of Social Behavior and Personality*, *15*(4), 413–430.

De Vries, R. E., Roe, R. A., & Taillieu, T. C. B. (1999). On charisma and need for leadership. *European Journal of Work and Organizational Psychology, 8*(1), 109–133.

De Vries, R. E., Roe, R. A., & Taillieu, T. C. B. (2002). Need for leadership as a moderator of the relationships between leadership and individual outcomes. *Leadership Quarterly*, *13*(2), 121–137.

De Vries, R. E., & Van Gelder, J. L. (2005). Leadership and need for leadership: Testing an implicit followership theory. In B. Schyns & J. R. Meindl (Eds.), *Implicit leadership theories: Essays and explorations*. (pp. 277–303). Greenwich, CT: Information Age.

Dinh, J. E., & Lord, R. G. (2012). Implications of disposition and process views of traits for individual difference research in leadership. *Leadership Quarterly, 23,* 651–669.

Dulebohn, J. H., Bommer, W. H., Liden, R. C., Brouer, R. L., & Ferris, G. R. (2012). A meta-analysis of antecedents and consequents of leader–member exchange: Integrating the past with an eye toward the future. *Journal of Management*, *38*(6), 1715–1759.

Dunning, D. (2003). The relation of self to social perceptions. In M. L. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 421–441). New York, NY: Guilford Press.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109,* 573–598.

Eden, D., & Leviatan, U. (1975). Implicit leadership theory as a determinant of the factor structure underlying supervisory behavior scales. *Journal of Applied Psychology*, *60*, 736–741.

Ellis, L. (1994). The high and the mighty among men and beast: How universal is the relationship between height (or body size) and social status. In L. Ellis (Ed.), *Social stratification and socioeconomic inequality: Reproductive and interpersonal aspects of dominance and status* (Vol. 2, pp. 93–111). Westport, CT: Praeger.

Engle, E. M., & Lord, R. G. (1997). Implicit leadership theories, self-schemas and leader–member exchange. *Academy of Management Journal, 40*(4), 988–1010.

Ensari, N., & Murphy, S. E. (2003). Cross-cultural variations in leadership perceptions and attributions of charisma to the leader. *Organizational Behavior and Human Decision Processes, 92*, 52–66.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013–1027.

Felfe, J., & Heinitz, K. (2010). The impact of consensus and agreement of leadership perceptions on commitment, OCB, and customer satisfaction. *European Journal of Work and Organizational Psychology, 19,* 279–303.

Felfe, J., & Schyns, B. (2010). Followers' personality and the perception of transformational leadership: Further evidence for the similarity hypothesis. *British Journal of Management, 21*(2), 393–410.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York, NY: Gilford Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). New York, NY: Academic Press.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition.* New York, NY: Random House.

Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture* (2nd ed.). New York, NY: McGraw-Hill.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117,* 39–66.

Forster, J., Higgins, E. T., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade off or separate strategic concerns? *Organizational Behavior and Human Performance, 90,* 148-164.

Foti, R. J., & Hauenstein, N. M. A. (2007). Patterns and variables: Evidence from leadership emergence and effectiveness. *Journal of Applied Psychology, 92,* 347–355.

Foti, R. J., & Lord, R. G. (1987). Prototypes and scripts: The effects of alternative methods of processing information. *Organizational Behavior and Human Decision Processes, 39,* 318–341.

Fowles, D. C. (1987). Application of a behavioral theory of motivation to the concepts of anxiety and impulsivity. *Journal of Research in Personality, 21,* 417–435.

Fowles, D. C. (1994). A motivation theory of psychopathology. In W. Spaulding (Ed.), *Nebraska Symposium on Motivation: Integrated views of motivation, cognition, and emotion* (Vol. 41) (pp. 181-238). Lincoln, NE: University of Nebraska Press.

Fuller, B. J., Patterson, C. E. P., Hester, K., & Stringer, D. Y. (1996). A quantitative review of research on charismatic leadership. *Psychological Reports*, *78*, 271–289.

Gara, M. A., Woolfolk, R. L., Cohen, B. D., Goldston, R. B., Allen, L. A., & Novalany, J. (1993). Perception of self and other in major depression. *Journal of Abnormal Psychology, 102,* 93–100.

Gawley, T., Perks, T., & Curtis, J. (2009). Height, gender, and authority status at work: Analysis for a national sample of Canadian workers. *Sex Roles, 60*(3), 208–222.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble with thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509–517.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54,* 733–739.

Graen. G. B. (1975). Role-making processes within complex organizations. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago, IL: Rand McNally.

Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. New York, NY: Oxford University Press.

Gray, J. A. (1985). Issues in the neuropsychology of anxiety. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 5–25). Hillsdale, NJ: Erlbaum.

Gray, J. A. (1987). *The psychology of fear and stress* (2nd ed.). Cambridge, UK: Cambridge University Press.

Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. (1974). Race and sex as determinates of ratings by potential employers in simulated work sampling task. *Journal of Applied Psychology, 59,* 705–711.

Hanges, P. J., Lord, R. G., & Dickson, M. W. (2000). An information-processing perspective on leadership and culture: A case for connectionist architecture. *Applied Psychology: An International Review, 49,* 133–161.

Hansbrough, T. K. (2012). The construction of a transformational leader: Follower attachment and leadership perceptions. *Journal of Applied Social Psychology, 42*(6), 1533–1549.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*(3), 258–268.

Hastie, R., & Pennington, N. (1989). Notes on the distinction between memory-based versus on-line judgments. In J. N. Bassili (Ed.), *Online cognition in person perception* (pp. 1–17). Hillsdale, NJ: Erlbaum.

Hazan, C., & Shaver, P. R. (1990). Love and work: An attachment–theoretical perspective. *Journal of Personality and Social Psychology*, *59*(2), 270–280.

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.

Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality*, *10*(6), 3–26.

Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology, 1,* 393–398.

Heilman, M. E., & Haynes, M. C. (2005). No credit where credit is due: Attributional rationalization of women's success in male–female teams. *Journal of Applied Psychology, 90,* 905–916.

Hensley, W. E. (1993). Height as a measure of success in academe. *Psychology: A Journal of Human Behavior, 30,* 40–46.

Higgins, E. T., Idson, L. C., Freitas, A. L., Spiegel, J., & Molden, D. C. (2003). Transfer of value from fit. *Journal of Personality and Social Psychology, 84,* 1140-1153.

Hoffman, E., & Lord, R. G. (2013). A taxonomy of event-level dimensions: Implications for understanding leadership processes, behavior, and performance. *Leadership Quarterly*, *24*(4), 558–571.

Hofmann, D. A., Morgeson, F. P., & Gerras, S. J. (2003). Climate as a moderator of the relationship between leader–member exchange and content specific citizenship: Safety climate as an exemplar. *Journal of Applied Psychology, 88,* 170–178.

Human, L. J., & Biesanz, J. C. (2011). Through the looking glass clearly: Accuracy and assumed similarity in well-adjusted individuals' first impressions. *Journal of Personality and Social Psychology*, *100*(2), 349–364.

Hunter, S. T., Bedell-Avers, K. E., & Mumford, M. D. (2007). The typical leadership study: Assumptions, implications, and potential remedies. *Leadership Quarterly*, *18*(5), 435–446.

Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition* (pp. 179–236). Hillsdale, NJ: Erlbaum.

Isen, A. M. (1993). Positive affect and decision making. In M. Lewis & J. Haviland (Eds.), *Handbook of emotions* (pp. 261–277). New York, NY: Guilford Press.

Isen, A. M., & Daubman, K. A. (1984). The influence of affect on categorization. *Journal of Personality and Social Psychology, 47,* 1206–1217.

Isen, A. M., & Means, B. (1983). The influence of positive affect on decision-making strategy. *Social Cognition, 2,* 18–31.

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perceptions of relationships in the environment. *Journal of Personality and Social Psychology, 78,* 943–955.

Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. (2008). The strong, sensitive type: Effects of stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes, 106,* 39–60.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220–266). New York, NY: Academic Press.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129,* 339–375.

Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology, 89,* 428–441.

Jung, D. I., Yammarino, F. J., & Lee, J. K. (2009). Moderating role of subordinates' attitudes on transformational leadership and effectiveness: A multi-cultural and multi-level perspective. *Leadership Quarterly, 20*(4), 586–603.

Kaiser, R. B., Hogan, R. C., & Craig, S. B. (2008). Leadership and the fate of organizations. *American Psychologist*, *63*(2), 96–110.

Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal, 38,* 1036–1051.

Kark, R., & Shamir, B. (2002). The dual effect of transformational leadership: Priming relational and collective selves and further effects on followers. In B. J. Avolio & F. J. Yammarino (Eds.), *Transformational and charismatic leadership: The road ahead* (Vol. 2, pp. 67–91). Oxford, UK: Elsevier Science.

Kark, R., & Van Dijk, D. (2007). Motivation to lead, motivation to follow: The role of self-regulatory focus in leadership processes. *Academy of Management Review, 32,* 500–528.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192–240). Lincoln, NE: University of Nebraska Press.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review, 19,* 195–229.

Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 379–392.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research, 52 (1),* 31-60.

Lee, D. M., & Alvares, K. M. (1977). Effects of sex on descriptions and evaluations of supervisory behavior in a simulated industrial setting. *Journal of Applied Psychology, 62,* 405–410.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology, 72,* 275–287.

Lewin, K. (1951). *Field theory in social science*. New York, NY: Harper & Row.

Lewter, J., & Lord, R. G. (1992). *Affect, self-schemas and transformational leadership.* Paper presented at the Academy of Management Conference, Las Vegas, NV.

Liberman, N., Troupe, Y., & Stephan, E. (2007). Psychological distance. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 353-383). New York, NY: Guilford Press.

Liden, R. C., Erdogan, B., & Bauer, T. N. (2006). *The early development of leader–member exchange: A longitudinal investigation.* Paper presented at the Academy of Management Conference, Atlanta, GA.

Liden, R. C., & Maslyn, J. M. (1998). Multidimensionality of leader–member exchange: An empirical assessment through scale development. *Journal of Management, 24,* 43–72.

Liden, R. C., Wayne, S. J., & Stilwell, D. (1993). A longitudinal study of the early development of leader–member exchanges. *Journal of Applied Psychology, 78*(4), 662–674.

Little, A. C. (2014). Facial appearance and leader choice in different contexts: Evidence for task

   contingent selection based on implicit and learned face–behaviour/face–ability associations.

   Leadership Quarterly, 25, 865–874.

Locke, V., MacLeod, C., & Walker, I. (1994). Automatic and controlled activation of

   stereotypes: Individual differences associated with prejudice. *British Journal of Social

   Psychology, 33,* 29–46.

London, M., & Poplawski, J. R. (1976). Effects of information on stereotype development in

   performance appraisal and interview contexts. *Journal of Applied Psychology, 61,* 199–

   205.

Lord, R. G. (1985). Accuracy in behavioral measurement: An alternative definition based on

   raters' cognitive schema and signal detection theory. *Journal of Applied Psychology, 70*,

   66–71.

Lord, R. G., & Brown, D. J. (2004). *Leadership processes and follower self-identity*. Mahwah,

   NJ: Erlbaum.

Lord, R. G., Brown, D. J., & Freiberg, S. J. (1999). Understanding the dynamics of leadership:

   The role of follower self-concepts in the leader/follower relationship. *Organizational

   Behavior and Human Decision Processes*, *78*(3), 167–203.

Lord, R. G., Brown, D. J., Harvey, J. L., & Hall, R. J. (2001). Contextual constraints on

   prototype generation and their multilevel consequences for leadership. *Leadership

   Quarterly*, *12*(3), 331–338.

Lord, R. G., & Maher, K. J. (1991). *Leadership and information processing: Linking perceptions and performance*. Boston, MA: Routledge.

Lowe, K. B., Kroeck, K. G., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the MLQ literature. *Leadership Quarterly*, *7*, 385–425.

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*(1), 37–47.

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, *102*(1), 72–90.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognitions, emotion, and motivation. *Psychological Review, 98,* 224–253.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38*, 299–337.

Martell, R. F. (1996). What mediates gender bias in work behavior ratings? *Sex Roles, 35*, 153–169.

Martell, R. F., & Evans, D. P. (2005). Source-monitoring training: Toward reducing rater expectancy effects in behavioral management. *Journal of Applied Psychology*, *90*(5), 956–963.

Martell, R. F., & Guzzo, R. A. (1991). The dynamics of implicit theories of group performance: When and how do they operate? *Organizational Behavior and Human Decision Processes, 50,* 51–74.

Martinko, M. J., & Gardner, W. L. (1987). The leader/member attribution process. *Academy of Management Review, 12,* 235–249.

Martinko, M. J., Moss, S. E., Douglas, S. C., & Borkowski, N. (2007). Anticipating the inevitable: When leader and member attribution styles clash. *Organizational Behavior and Human Decision Processes, 104,* 158–174.

Mayer, J. D., Mamberg, M. H., & Volanth, A. J. (1988). Cognitive domains of the mood system. *Journal of Personality, 56,* 453–486.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20,* 709–734.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex. *Psychological Review*, *102*(3), 419–457.

McElroy, J. C., & Hunger, J. D. (1988). Leadership theory as a causal attribution of performance. In J. G. Hunt, B. R. Baliga, H. P. Dachler, & C. A. Schriesheim (Eds.), *Emerging leadership vistas* (pp. 169–183). Lexington, MA: Lexington.

Meindl, J. R. (1990). On leadership: An alternative to the conventional wisdom. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (12th ed., pp. 159–203). Greenwich, CT: JAI Press.

Meindl, J. R., & Ehrlich, S. B. (1988). Developing a "romance of leadership" scale. *Proceedings of the Eastern Academy of Management,* 133–135.

Meindl, J. R., Ehrlich, S. B., & Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly, 30,* 78–102.

Mikulincer, M., Gillath, O., & Shaver, P. R. (2002). Activation of the attachment systems in adulthood: Threat related primes increase the accessibility of mental representations of attachment figures. *Journal of Personality and Social Psychology*, *83*(4), 881–895.

Mikulincer, M., Shaver, P. R., & Pereg, D. (2003). Attachment theory and affect regulation: The dynamics, development, and cognitive consequences of attachment-related strategies. *Motivation and Emotion*, *27*, 77–102.

Morgeson, F. P. (2005). The external leadership of self-managing teams: Intervening in the context of novel and disrupting events. *Journal of Applied Psychology*, *90*(3), 497–508.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67,* 949–971.

Mount, M. K., & Scullen, S. E. (2001). Multisource feedback ratings: What do they really measure? In M. London (Ed.), *How people evaluate others in organizations* (pp. 155-176). Mahwah, NJ: Lawrence Erlbaum Associates.

Naidoo, L. J., Kohari, N. E., Lord, R. G., & DuBois, D. A. (2010). "Seeing" is retrieving: Recovering emotional content in leadership ratings through visualization. *Leadership Quarterly*, *21*(5), 886–900.

Norenzayan, A., Choi, I., & Nisbett, R. E. (1999). Eastern and Western perceptions of causality

    for social behavior: Lay theories about personality and social situations. In D. Prentice &

    D. Miller (Eds.), *Cultural divides: Understanding and overcoming group conflict* (pp.

    239–273). New York, NY: Sage.

O'Leary, V. E. (1974). Some attitudinal barriers to occupational aspirations in women.

    Psychological Bulletin, 81, 809–826.

Olivola, C. Y., Eubanks, D. L., & Lovelace, J. B. (2014). The many (distinctive) faces of

    leadership: Inferring leadership domain from facial appearance. *Leadership Quarterly, 25*,

    817–834.

Operario, D., & Fiske, S. T. (2001). Causes and consequences of stereotypes in organizations. In

    M. London (Ed.), *How people evaluate others in organizations* (pp. 45-62). Mahwah, NJ:

    Lawrence Erlbaum Associates.

Orne, M. T. (2009). Demand characteristics and the concept of quasi-controls. In R. Rosenthal &

    R. L. Rosnow (Eds.), Artifacts in behavioral research (pp. 110-137). New York: Oxford

    University Press.

Pfeffer, J. (1978). The ambiguity of leadership. In M.W. McCall & M. M. Lombardo (Eds.),

    Leadership: Where else can we go? (pp. 13–34). Durham, NC: Duke University Press.

Phillips, J. S. (1984). The accuracy of leadership ratings: A cognitive categorization perspective.

    Organizational Behavior and Human Performance, *33*(1), 125–138.

Phillips, J. S., & Lord, R. G. (1981). Causal attributions and perceptions of leadership.

    Organizational Behavior and Human Performance, 28, 143–163.

Phillips, J. S., & Lord, R. G. (1982). Schematic information processing and perceptions of

    leadership in problem-solving groups. *Journal of Applied Psychology*, *67*(4), 486–492.

Popper, M. (2013). Leaders perceived as distant and close: Some implications for psychological

    theory on leadership. *Leadership Quarterly*, *24*(1), 1–8.

Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and managerial stereotypes:

    Have the times changed? *Journal of Management, 28,* 177–193.

Riggio, R. E., Chaleff, I., & Lipman-Blumen, J. (Eds.). (2008). *The art of followership: How*

    great followers create great leaders and organizations. San Francisco, CA: Jossey-Bass.

Rosette, A. S., Leonardelli, G. J., & Phillips, K. W. (2008). The white standard: Racial bias in

    leader categorization. *Journal of Applied Psychology, 93,* 758–777.

Rush, M. C., Phillips, J. S., & Lord, R. G. (1981). The effects of a temporal delay in ratings

    leader behavior descriptions: A laboratory investigation. *Journal of Applied Psychology,*

    66, 442-450.

Rush, M. C., Thomas, J. C., & Lord, R. G. (1977). Implicit leadership theory: A potential threat

    to the internal validity of leader behavior questionnaires. *Organizational Behavior and*

    Human Performance, *20*, 93–110.

Russell, D. W. (1991). The measurement of attribution process: Trait and situational approaches.

    In S. L. Zelden (Ed.), *New models, new extensions of attribution theory* (pp. 55-83). New

    York, NY: Springer-Verlag.

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins & R. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 527–561). New York, NY: Guilford Press.

Schyns, B., Felfe, J., & Blank, H. (2007). Is charisma hyper-romanticism? Empirical evidence from new data and a meta-analysis. *Applied Psychology: An International Review, 56,* 505–527.

Schyns, B., Kroon, B., & Moors, G. (2008). Follower characteristics and the perception of leader–member exchange. *Journal of Managerial Psychology*, *23*(7), 772–788.

Scott, K. A., & Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organizational Behavior and Human Decision Processes*, *101*(2), 230–242.

Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85,* 956–970.

Shamir, B. (1995). Social distance and charisma: Theoretical notes and an exploratory study. *Leadership Quarterly, 6*(1), 19–48.

Shamir, B. (2007). Introduction: From passive recipients to active co-producers: The roles of followers in the leadership process. In B. Shamir, R. Pillai, M. C. Bligh, & M. Uhl-Bien (Eds.), *Follower-centered perspectives on leadership: A tribute to James R. Meindl*. Greenwich, CT: Information Age.

Shamir, B., Pillai, R., Bligh, M. C., & Uhl-Bien, M. (Eds.). (2006). *Follower-centered perspectives on leadership: A tribute to the memory of James R. Meindl*. Greenwich, CT: Information Age.

Shondrick, S. J., Dinh, J. E., & Lord, R. G. (2010). Developments in implicit leadership theory and cognitive science: Applications to improving measurement and understanding alternatives to hierarchical leadership. *Leadership Quarterly*, *21*(6), 959–978.

Shore, L. M., & Thornton, G. C. (1986). Effects of gender on self- and supervisory ratings. *Academy of Management Journal, 29,* 115–129.

Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 391–445). New York, NY: McGraw-Hill.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108–131.

Smith, J. A., & Foti, R. J. (1998). A pattern approach to the study of leader emergence. *Leadership Quarterly, 9,* 147-160.

Srull, T. K., & Wyer, R. S. (1989). Person memory and person judgment. *Psychological Review, 96*(1), 58–83.

Staw, B. M. (1975). Attribution of the "causes" of performance: A general alternative interpretation of cross-sectional research on organizations. *Organizational Behavior and Human Performance, 13,* 414–432.

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2013). Tall claims? Sense and nonsense about the importance of height of US presidents. *Leadership Quarterly, 24,* 159–171.

Sulsky, L. M., & Day, D. V. (1994). Effects of frame of reference training on rater accuracy under alternative time delays. *Journal of Applied Psychology, 79,* 535–543.

Sy, T., Shore, L. M., Strauss, J., Shore, T. H., Tram, S., Whiteley, P., & Ikeda-Muromachi, K. (2010). Leadership perceptions as a function of race–occupation fit: The case of Asian Americans. *Journal of Applied Psychology, 95,* 902–919.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues, 25,* 79–97.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum.

Terborg, J. R., & Ilgen, D. R. (1975). A theoretical approach to sex discrimination in traditionally masculine occupations. *Organizational Behavior and Human Performance, 13,* 352–376.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308,* 1623–1626.

Trichas, S. & Schyns, B. (2012). The face of leadership: Perceiving leaders from facial expression. *Leadership Quarterly 23*, 545–566.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385-398.

Turban, D. B., & Jones, A. P. (1988). Supervisor–subordinate similarity: Types, effects, and mechanisms. *Journal of Applied Psychology*, *73*(2), 228–234.

Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., & Johnson, R. E. (2012). Getting explicit about the implicit: A taxonomy of implicit measures and guide for their use in organizational research. *Organizational Research Methods*, *15*, 553–601.

Verduyn, P., & Brans, K. (2012). The relationship between extraversion, neuroticism and aspects of trait affect. *Personality and Individual Differences, 52,* 664–669.

Watson, D. (1988). Intraindividual and interindividual analyses of positive and negative affect: Their relation to health complaints, perceived stress, and daily activities. *Journal of Personality and Social Psychology, 54,* 1020–1030.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465–490.

Watson, D., & Clark, L. A. (1992). Affect, personality and social activity. *Journal of Personality and Social Psychology, 63,* 1011–1025.

Watson, D., Clark, L. A., & Carey, G., (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97,* 346–353.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*(3), 546–558.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820–838.

Wayne, S. J., & Ferris, G. R. (1990). Influence tactics, affect, and exchange quality in supervisor–subordinate interactions: A laboratory experiment and field study. *Journal of Applied Psychology*, *75*(5), 487–499.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. A. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). San Diego, CA: Academic Press.

Weiss, H. M., & Adler, S. (1981). Cognitive complexity and the structure of implicit leadership theories. *Journal of Applied Psychology*, *66*, 69–78.

Welsch, G. S. (1975). *Creativity and intelligence: A personality approach.* Chapel Hill, NC: Institute of Research in Social Science.

Werner, D. (1982). Chiefs and presidents: A comparison of leadership traits in the United States and among the Mekranoti-Kayapo of central Brazil. *Ethos, 10*(2), 136–148.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 71,* 262–274.

Wright, J., & Mischel, W. (1982). Influence of affect on cognitive social learning person variables. *Journal of Personality and Social Psychology, 43,* 901–914.

Yammarino, F. J., Dubinsky, A. J., & Spangler, W. D. (1998). Transformational and contingent

　　　reward leadership: Individual, dyad and group levels of analysis. *Leadership Quarterly*,

　　　9(1), 27–54.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American*

　　　Psychologist, 35(2), 151–175.