

**EFFECTS OF PERSONALITY AND GENDER ON SELF-OTHER AGREEMENT
IN RATINGS OF LEADERSHIP**

ROB AUSTIN MCKEE
Assistant Professor, Organizational Behavior
Marylin Davis College of Business
University of Houston--Downtown
mckeer@uhd.edu

YIH-TEEN LEE
Associate Professor of Managing People in Organizations
Department of Managing People in Organizations
IESE Business School
ylee@iese.edu

LEANNE E. ATWATER
Professor of Management
C.T. Bauer College of Business
University of Houston
leatwater@uh.edu

JOHN ANTONAKIS
Professor of Organizational Behavior
Faculty of Business and Economics
University of Lausanne
john.antonakis@unil.ch

In press
February 2018

Journal of Occupational and Organizational Psychology

Effects of Personality and Gender on Self-Other Agreement in Ratings of Leadership

Abstract

We explore the role of leader personality (i.e., the Big 5 traits Conscientiousness, Agreeableness, Openness, Extraversion, and Neuroticism) and gender in self-other (dis)agreement (SOA) in ratings of leadership. We contend that certain aspects of the leader's persona may be more or less related to self- or other-ratings of the leader's behavior if those aspects are (a) more or less observable by others, (b) more or less related to internal thoughts vs. external behaviors, (c) more or less prone to self-enhancement or self-denigrating biases, or (d) more or less socially desirable. We utilize statistical methodologies that capture fully the effects of multiple independent variables on the congruence between two dependent variables (Edwards, 1995), which previously have not been applied to this area of research. Our results support hypotheses predicting less SOA as leader Conscientiousness increases and greater SOA as Agreeableness and Neuroticism increase. Additionally, we found gender to be an important factor in SOA; female leaders exhibited greater SOA than did their male counterparts. We discuss the implications of these findings, limitations, and future research directions.

Keywords: self-other agreement; instrumental leadership; personality; gender

Practitioner Points:

- Popular practices such as 360-degree feedback may reveal discrepancies between a person's self-ratings and other's ratings.
- Though often attributed to a lack of self-awareness, these discrepancies also may be explained by factors such as the personality and gender of the focal individual.

Effects of Personality and Gender on Self-Other Agreement in Ratings of Leadership

Self-other (dis)agreement (SOA) in behavior ratings between leaders and their subordinates, peers, and superiors has been studied as a concept relevant to individual and organizational outcomes for over 25 years (Fleenor, Smither, Atwater, Braddy, & Sturm, 2010)¹. It is an important construct for both scholars and practitioners given the rising popularity of multi-rater feedback and assessment systems in organizations. Much of the interest in SOA “derives from two primary factors: (a) it is posited to be an indicator of self-awareness and (b) it appears to be related to several outcomes of interest including leader effectiveness and derailment” (Fleenor et al., 2010, p. 1005). Greater discrepancies in ratings between the self and others presumably indicate lesser self-awareness, and lesser self-awareness can underlie a number of problems, particularly for leaders who are unaware of their strengths and weaknesses and oblivious to how others see them.

In this study, we seek to answer the following question: Do individual characteristics of leaders contribute to SOA? This question has both theoretical and practical relevance. A comprehensive theoretical model of SOA proposed by Atwater and Yammarino (1997) suggested a leader’s personality traits and gender were important factors contributing to both self-and other-ratings. However, with few exceptions, little work has been done to explore leader characteristics and how they might affect self-and other-ratings and ultimately agreement on those ratings. This study addresses these issues thereby advancing the theory surrounding antecedents of SOA. Theoretically, our research proposes that rather than merely affecting how a leader rates himself or herself, his or her traits may ultimately affect agreement because of the

¹ As will be explained in detail below, in this article we conceptualize SOA as the degree of agreement between self- and other-rating of a specific assessment of a focal leader; we applied the method of multivariate regression (Edwards, 1995) to examine its antecedent, instead of treating SOA as correlations between self- and other-ratings.

differential ways in which traits influence self, versus other ratings. Specifically, we propose and test the idea that SOA in leadership assessment (or other types of appraisals) may be influenced by factors associated with individual differences of the self-rater stemming primarily from the Big Five personality traits. We suggest that SOA may be, in part, a reflection of personality traits of the self-rater and less of a reflection of self-awareness per se. If this suggestion is demonstrated to be the case, it is of both theoretical and practical relevance because it will call into question the widely accepted practice of using SOA as a proxy for self-awareness. For example, if SOA can in part be explained by the Extraversion or Agreeableness of the self-rater, the results would suggest that SOA is a less valid indicator of self-awareness, or that personality of the self-rater must at least be accounted for if one wants to use it as a measure of self-awareness.

This study also raises an important practical question because practitioners worldwide are providing 360-degree feedback to managers as part of leadership development initiatives. The rationale is that providing leaders with information about the degree to which their self-ratings agree with ratings from other groups will enlighten them about others' perceptions of their performance, hopefully engendering greater performance via increased self-awareness. If, however, the degree of agreement is mainly an artifact of the leader's characteristics (e.g., personality), then it may not be a useful construct for leader development beyond understanding one's traits. For example, the hope in using 360-degree feedback is that it will inform leaders about their strengths and weaknesses, which simultaneously should reduce self-other disagreement (i.e., increase self-awareness) and lead to greater leader effectiveness. But, if being an under or over-estimator relative to others' ratings is largely about personality traits (e.g., introversion or extraversion), then feedback of self-and other-ratings designed to decrease self-

other discrepancies may not be fruitful. Specifically, the Big-Five personality traits and gender, among focal leaders (i.e., self-raters) may predict not only self-ratings but also other-ratings and thereby contribute to self-other rating disagreement.

BACKGROUND

The attention to SOA in work settings has been addressed in topics such as (dis)agreement between the self and various rating sources (e.g., subordinate, peer, superior; see Harris & Schaubroeck, 1988), as well as factors such as demographics that predict discrepancies between self and various rater groups (Brutus, Fleenor, & McCauley, 1999). Outcomes of SOA studied in the realm of leadership include leader performance, derailment, and follower attitudes (Atwater & Brett, 2005; Atwater & Yammarino, 1992; McCall & Lombardo, 1983). Antecedents of SOA such as self-rater characteristics (e.g., age and gender) have also been studied. Fleenor et al. (2010) provided a summary of studies conducted in the area of SOA. To generalize from these studies, some conclusions are that: degree of agreement between self- and other-raters is generally lower than agreement between other rating sources with each other (Harris & Schaubroeck, 1988); overrating oneself relative to others' ratings is associated with worse outcomes than is underrating oneself (Bass & Yammarino, 1991; Van Velsor, Taylor, & Leslie, 1993); males and individuals with longer organizational tenure are likely to overrate themselves relative to others (Brutus et al., 1999); and, males provide higher self-ratings than do females (Paustian-Underdahl, Walker, & Woehr, 2014). Additionally, studies of SOA have advanced methodologically. Whereas difference scores, categories of agreement based on difference scores, and correlations seem to have been used in the earlier years, research in SOA nowadays relies on more sophisticated and appropriate approaches such as polynomial regression and

multivariate regression (Edwards, 1994; 1995; Edwards & Parry, 1993) as the preferred approaches.

Self- and Other-Ratings

Numerous studies dating back as far as the 1980s have shown that self- and other-ratings of a focal person's behavior largely disagree (Harris & Schaubroeck, 1988; Mabe & West, 1982). For instance, Furnham and Stringfield (1998) found that of the 20 behaviors rated by self and others on a developmental 360-degree leadership feedback tool, self-ratings were higher than ratings from any other rater groups on all 20 behaviors assessed. It is important to note here that in this paper we are not concerned with the accuracy of self- or other-ratings, but simply on self-other (dis)agreement in terms of convergence and divergence of those ratings from one another. Indeed, per a review conducted by Fleenor et al. (2010), neither self- nor other-ratings are presumed to be more accurate; congruence is the primary concern. There are numerous reasons self- and other-ratings may converge or diverge.

One avenue by which self-ratings may diverge from other-ratings is built upon the concept of observability. Some aspects of a person (e.g., behaviors, performance, personality) are more or less observable depending on the party of concern (i.e., who is being asked to rate those aspects). Thus, certain aspects of a person may be best judged by the self and certain other aspects may be best judged by others (Vazire, 2010). According to this model of self-other knowledge asymmetry, "the self should be more accurate than others for traits low in observability (e.g., neuroticism), whereas others should be more accurate than the self for traits high in evaluativeness (e.g., intellect)" (Vazire, 2010, p. 281). Within the leader-follower relationship, followers are not likely to observe some of the activities involved in planning and goal-setting, and may not be aware of the (full spectrum of) objective performance measures to

which the leader is held accountable. Conversely, followers may be positioned especially well to evaluate the leader's degree of consideration and communication skills. As noted by Smither, Brett, and Atwater (2007, p. 204), "in the absence of direct feedback, managers are likely to have difficulty knowing whether they are viewed by others as high in consideration (e.g., warm, friendly, trustworthy, supportive, approachable). In contrast, even in the absence of direct feedback, managers might be able to accurately evaluate the extent to which they are high in initiating structure (e.g., whether they frequently schedule work, assign tasks, specify procedures, clarify expectations) because these behaviors are somewhat more objective than behaviors related to consideration (e.g., being seen as 'supportive' or 'warm')." It is important to note that observability may depend on situational factors, meaning that some traits or behaviors may be more or less observable depending on the environment (de Vries, Realo, & Allik, 2016; Rauthmann, 2012; Tett & Burnett, 2003).

Another possible explanation for self-other rating discrepancy has to do with the "lens" through which the self and others perceive and interpret the focal person's behavior. Pronin (2008, p. 1177) aptly describes this phenomenon, stating, "differences in how people see themselves versus others are systematic and predictable, and rooted in basic processes of human perception." Pronin (2008, p. 1177) asserts that "we tend to observe ourselves via 'introspection' (looking inwards to thoughts, feelings, and intentions) and others via 'extrospection' (looking outwards to observable behavior)," because we have essentially proprietary access to our thoughts, feelings, and intentions, but cannot actually literally see (i.e., visually) ourselves as we can others, and as others can us. The Realistic Accuracy Model (RAM) (Funder, 2012) provides a similar perspective suggesting that personality judgments are more accurate when relevant behavioral information is available, detectable, and interpreted correctly. We can add to this

explanation the fact that the self often has more instances on which to rate his or her behavior than do external observers. The result is that people's self-impressions and their impressions of others (or others' impressions of them) are often based on dissimilar or incomplete information.

Self-ratings may also diverge from other-ratings as a result of cognitive biases. Pronin (2008) describes the underpinnings of this tendency in terms of generally inflated self-views, inflated estimates of how much we think we know about others, misperceptions of others' thoughts and motives, and miscommunication. Several researchers have echoed this stance, arguing that self-ratings are distorted by leniency or self-enhancement bias and validity issues (Podsakoff & Organ, 1986; Thornton, 1980). Thus, although the individual should have more insights and instances on which to assess his or her behaviors than do observers, and thus we may wish to favor the validity of self-ratings over other-ratings, self-ratings are notoriously prone to self-enhancement and sometimes even self-denigrating biases, even when self-raters are informed of the various biases that may be affecting them (Dunning, Meyerowitz, & Holzberg, 1989; Pronin, Lin, & Ross, 2002). As noted in a recent review of SOA, "self-ratings alone, in general, are not considered to be accurate predictors of leadership because they are likely inflated by leniency bias" (Fleenor et al., 2010, p. 1006). As we will argue below, these self-enhancing or denigrating tendencies are influenced by personality traits of the self-rater and thus are likely to influence self-other agreement.

One more reason self-ratings may diverge from other-ratings is the tendency for socially-desirable responding. For instance, as discussed by Digman (1997) and in further detail below, socialization theories contend that the positive and often sizable correlations observed by certain dimensions of personality (e.g., Conscientiousness, Agreeableness, and Emotional Stability) may be driven by socially desirable responding, or by the underlying nature of the traits as socially

desirable. This rationale can be extended to argue that such traits are a manifestation of the socialization process itself. In other words, such traits represent how we learn to act socially appropriately with respect to what is and what is not socially desirable is multifaceted. Thus, although certain traits (e.g., Conscientiousness and Agreeableness) may be related (or at least correlated), they are distinct constructs representing different behavioral proclivities, and it is reasonable to assert that high levels of one trait (i.e., Conscientiousness) may cause one to inflate self-ratings of some capacity or behavior whereas high levels of the other trait (i.e., Agreeableness) may cause one to deflate or at least constrain self-ratings of the same behavior or capacity.

We expect that aspects of the leader's personality and gender influence four mechanisms, explained below, causing (in)congruence between self- and other-ratings described immediately above, and thereby influence self-other agreement. As such, we propose to significantly expand our understanding of traits and their relationships to (dis)agreement by looking at the extent to which various leader traits predict self- and other-ratings of leadership, as well as the form of the relationships between self- and other-ratings in light of these traits.

Personality Traits as Predictors of Self and Other-ratings

Personality traits have been investigated as antecedents of a vast array of behaviors, such as performance (Barrick, Mount, & Judge, 2001; Salgado, 2003; Tett, Jackson, Rothstein, & Reddon, 1994), leadership (Bono & Judge, 2004; Cavazotte, Moreno, & Hickmann, 2012; Lord, de Vader, & Alliger, 1986), and counter-productive work behaviors (Mount, Ilies, & Johnson, 2006). The Big 5 personality traits—Conscientiousness, Agreeableness, Openness, Extraversion, and Neuroticism—have been the subject of many of these studies (Hurtz & Donovan, 2000; Judge, Higgins, Thoresen, & Barrick, 2006; Salgado, 2003).

Although the literature has addressed the extent to which the rater's personality traits influence the way he or she rates others (Ogunfowora, Bourdage, & Lee, 2010), the target's traits have rarely been looked at in terms of how they influence self-ratings or those provided by others (i.e., observers). Goffin and Anderson (2007) looked at self-rater personality as it relates to SOA in performance ratings, finding that self-superior rating differences were related to target self-esteem, achievement, and anxiety, but these effects were due to the traits' relationships to the self-ratings rather than the ratings provided by superiors. Sinha, Mesmer-Magnus, and Viswesvaran (2012) also studied traits and their relationships to SOA in performance ratings, finding that the traits related to disagreement due to distortions in self-ratings rather than other-ratings. Brutus et al. (1999) did not study the Big 5 but found that target empathy was positively related to self and other-ratings of leadership.

We argue that aspects of the leader's personality and gender influence SOA through at least four mechanisms:

1. Some personality traits are more or less observable, or lead to more or less observable behaviors, and thus likely have different relationships to self- and other-ratings due to self-other knowledge asymmetry (Vazire, 2010).

2. Factors related to introspection and extrospection will sway SOA in line with Funder's (1980, p. 473) statement: "As would be predicted from attribution research, subjects tend to rate themselves higher than do their peers on traits pertaining to inner states (e.g., 'is introspective'), while peers tend to rate them higher on traits pertaining to behaviors especially salient to an external observer (e.g., 'is personally charming')."

3. Certain traits will exacerbate self-enhancement or self-denigrating biases, making discrepancies in SOA more likely as a function of those traits (Kwan, John, Robins, & Kuang, 2008; Schaefer, Williams, Goodie, & Campbell, 2004).

4. Certain traits may compel people to act (or to perceive themselves as acting) in socially desirable ways, or such traits may be a manifestation of learning to act in socially desirable ways (Digman, 1997). Thus, a person high in one trait will want to be seen as embodying the socially desirable aspects of that trait. If we view social desirability as a property of certain traits, it may influence the relationship between such traits and SOA (Funder & Colvin, 1988; John & Robins, 1993).

As documented by previous research (Mount & Barrick, 1995; Tett, Jackson, & Rothstein, 1991), theory-driven personality-criterion relationships are more likely to result in important findings than are exploratory studies that do not propose specific relationships. As such, we explain how the personality traits relate to self- and other-ratings of leadership. The criterion we chose is instrumental leadership, which is a complement to a well-known leader style transformational leadership. Although the latter has been widely studied, Antonakis and House (2014) found that instrumental leadership (i.e., the use of leader expertise to adapt the organization to its environment and help followers succeed) was more strongly related to leader effectiveness and similarly related to employee satisfaction. Additionally, because instrumental leadership is behaviorally oriented, we believe it will be less subject to halo biases than the more emotion-laden and positively-valenced transformational leader constructs. However, our focus is on SOA rather than any specific leadership theory. We did gather data on transformational leadership too, though we are only reporting the instrumental leadership results given the much

stronger psychometric validity of the scales². We focus on instrumental leadership *in toto* and theorize how personality may explain SOA on instrumental leadership per se; however, as we explain later (based on empirical justifications), we report more fine-grained results on the subcomponents of the instrumental leadership scales.

Instrumental Leadership

Instrumental leadership (IL) is conceptualized as “the application of leader expert knowledge on monitoring of the environment and of performance, and the implementation of strategy and tactical solutions” (Antonakis & House, 2014, p. 749); it is seen as highly prototypical of effective leadership—on par with transformational and contingent reward leadership—and strongly predicts effectiveness outcomes. As discussed by Antonakis and House (2002, 2014), IL comprises four subcomponents: environmental monitoring (EM), strategy formulation and implementation (SF), path-goal facilitation (PG), and outcome monitoring (OM). EM refers to the leader scanning the internal and external organizational environments to determine organizational strengths and weaknesses and to identify opportunities and threats. SF refers to the leader developing policies, goals, and objectives that support the strategic vision and mission of the organization. PG refers to the leader providing followers direction, support, and resources, removing obstacles to their achieving goals, and providing them path-goal clarifications. OM refers to the leader providing followers performance-enhancing feedback to help them achieve goals. As argued by Antonakis and House (2002, 2014), IL is intended to complement the transactional and transformational components of the full-range leadership paradigm (e.g., Bass, 1985). In that regard, IL is a qualitatively different style of leadership as

² Using appropriate criteria to judge the fit of the transformational leadership model showed that its factor structure was not tenable. Still, for comparative purposes, we do report detailed results on this construct too in the supplementary material.

compared to transactional and transformational leadership because IL is wholly focused on strategic and task-oriented leadership functions; as such, it does not include the contingent rewards and sanctions underlying transactional leadership, nor does it include the affective and value-based aspects or inspirational appeals inherent to transformational leadership.

HYPOTHESIS DEVELOPMENT

In a general way, our hypotheses contend that various aspects of a leader's personality (i.e., Conscientiousness, Agreeableness, Openness, Extraversion, and Neuroticism) are differentially related to self- and other-ratings of a target's IL behaviors. More observable traits will be more strongly related to others' ratings, and less observable traits (more internally manifested) will be more strongly related to self-ratings. Also, traits more susceptible to self-enhancement will likely result in higher self-ratings than other-ratings. In other words, if a trait is related to leadership and that trait is more observable and less influenced by self-enhancement, it will be more likely related to others' ratings. However, if the trait is related to leadership and less observable but more influenced by self-enhancement, it will be more likely related to self-ratings of leadership. These self and other relationship differences will thereby affect the degree of SOA in those ratings. In order to craft our hypotheses, we must predict whether leaders' increasing expression of a given dimension of personality (e.g., Conscientiousness) will primarily affect self- or other-ratings of their leadership and whether that effect will be in a positive or negative direction (e.g., higher Conscientiousness is related to higher instrumental leadership; higher Neuroticism is related to lower instrumental leadership).

However, to infer whether a personality trait enhances or reduces SOA, we must predict the relative magnitudes of the self- and other-rating lines (i.e., whether people tend to overrate or underrate themselves) for specific leadership behaviors. As a starting point, self-ratings of

behaviors and traits provided by psychologically “healthy” persons generally are higher than others’ ratings due to self-enhancement (i.e., a self-serving bias) and socially-desirable responding (Digman, 1997; Dunning et al., 1989; John & Robins, 1994; Kwan, John, Kenny, Bond, & Robins, 2004; Paulhus & Reid, 1991; Pronin et al., 2002; Sedikides & Gregg, 2008; Taylor, 1989). The exception to this tendency is provided by depressives and others with low self-esteem (e.g., those high on Neuroticism; see Campbell & Fehr, 1990; Noles, Cash, & Winstead, 1985) who “may not only self-enhance less but may actually see themselves more negatively than they are seen by others” (John & Robins, 1994, p. 209). Therefore, as a general rule, we expect self-ratings to be higher than other-ratings for a prototypically effective leadership style like IL. The rationale for how the specific dimensions of personality will play a role in SOA, as well as exceptions to this general rule, will be discussed in detail below.

As an example, if we generally expect self-ratings to be higher than other ratings when considering most traits and behaviors (Dunning et al., 1989; Pronin et al., 2002) and we expect Conscientiousness to be more positively related to self-ratings than to other-ratings, then increasing levels of Conscientiousness among leaders should result in decreasing SOA. That is, self-ratings of IL will be both higher than other-ratings and more positively sloped (i.e., more highly correlated with Conscientiousness) than other ratings, thereby increasing the discrepancy between self- and other-ratings as Conscientiousness increases. We expand upon this rationale immediately below.

Conscientiousness and Self- and Other-Ratings of Leadership

Primarily an aspect of character, Conscientiousness is proactive, shown via a need for achievement and commitment to work; it is also inhibitive, exhibited through moral scrupulousness and cautiousness (Costa, McCrae, & Dye, 1991). The Five-Factor Model (FFM;

see Salgado, 2003) facets of Conscientiousness include Competence, Order, Dutifulness, Achievement Striving, Self-Discipline, and Deliberation (Costa et al., 1991). Conscientiousness is a consistent predictor of leadership and job performance outcomes (Barrick et al., 2001; Hurtz & Donovan, 2000). Similar to other types of leadership shown to be positively related to Conscientiousness, we expect IL to be positively related to Conscientiousness.

The effect of Conscientiousness depends on several of the mechanisms mentioned above that may cause a divergence in SOA. For instance, Conscientiousness, as measured by the NEO PI-3 (McCrae, Costa, Jr., & Martin, 2005), has been assessed as moderately observable (lower than Extraversion but higher than Neuroticism, Agreeableness, and Openness to Experience) (de Vries et al., 2016). However, although observers may be able to see some behaviors that reflect Conscientiousness, such behaviors are largely internal and reflective (e.g., achievement striving, self-discipline, deliberation; Barrick, Patton, & Haugland, 2000; Connolly, Kavanagh, & Viswesvaran, 2007a). Moreover, Conscientiousness is a highly socially desirable trait (Digman, 1997; McFarland & Ryan, 2000; Stöber, 2001), and according to RAM (Funder, 1995), this feature means that Conscientiousness is susceptible to self-enhancement in self-ratings. Thus, there may be an increased tendency for conscientious individuals to engage in self-enhancement or even self-deception (Lee & Klein, 2002; Martocchio & Judge, 1997; Stöber, Dette, & Musch, 2002). Supporting this notion, Judge, LePine, and Rich (2006) found Conscientiousness to be positively related to self-ratings of leadership. Similarly, Visser, Ashton, and Vernon (2008) found Conscientiousness to be positively related to self-estimated ability. Cheng, Hui, and Cascio (2017) also found that conscientious individuals tend to make generous self-assessments. Judge et al. (2006) speculated about the psychological processes underlying this phenomenon; unlike narcissists, who may self-deceive due to a sense of grandiosity or true arrogance

(Emmons, 1987), conscientious individuals may do so for more defensive reasons, emanating from their need to maintain a positive self-image as a competent person (Burriss & Navara, 2002) and due to the importance they ascribe to their work (Digman, 1997). Cheng et al. (2017) cited the high achievement need of conscientious individuals as a driving force for self-enhancement bias.

Thus, the preponderance of the evidence seems to indicate a significant positive effect on self-ratings of IL as Conscientiousness increases, but little or no similar effect on other-ratings. Given the asymmetry in the relationship between self- and other-ratings and Conscientiousness, we also predict that Conscientiousness will contribute to self-other (dis)agreement such that disagreement will be greatest for leaders high in Conscientiousness. In other words, as leader Conscientiousness increases, so too will the discrepancy between self-ratings and other-ratings, because the concomitant increase in self-ratings of positive leadership behaviors will be greater than any changes in other-ratings of those same behaviors.

H1 Leader Conscientiousness will be more positively related to self-ratings than to other-ratings of IL.

H2 Self-other-ratings of IL will exhibit greater agreement at *lower* levels of leader Conscientiousness than at *higher* levels of Conscientiousness.

Agreeableness and Self- and Other-Ratings of Leadership

Agreeableness is primarily a dimension of interpersonal behavior and interaction, though it also influences self-image and social attitudes (Costa et al., 1991). In terms of interpersonal interaction, Agreeableness exists “along a continuum from compassion to antagonism” (Costa & McCrae, 1985, p. 2). The FFM facets of Agreeableness include Trust, Straightforwardness, Altruism, Compliance, Modesty, and Tender-Mindedness (Costa et al., 1991).

Agreeableness is related to several mechanisms that may cause a divergence in SOA. For instance, some have contended that expressions of Agreeableness may be low in observability, especially at low levels of the trait because people infrequently act highly disagreeably (de Vries et al., 2016). Others have argued that Agreeableness generally should be externally observable (Barrick et al., 2000; Carlson, Vazire, & Furr, 2011). Indeed, within the realm of leadership, many components of Agreeableness (e.g., straightforwardness, cooperation, modesty, sympathy) are observable to others, perhaps especially followers, because those components directly affect how the leader treats them. Because Agreeableness is characterized by such positive interpersonal behaviors that are likely observable at least to some degree, we expect others' ratings of a leader's IL may increase as Agreeableness increases.

Like Conscientiousness, the effects of socially-desirable responding on SOA also may be relevant to Agreeableness. Whereas Agreeableness and Conscientiousness are positively correlated traits, we propose divergent effects for the two traits with regard to self-ratings. Digman (1997) discusses broad causes responsible for the positive and often sizable correlations between Agreeableness, Conscientiousness, and Emotional Stability (i.e., the antipode of Neuroticism). Socialization theories contend that such correlations might be driven by socially desirable responding, or by the underlying nature of the traits as socially desirable. This rationale can be extended to argue that this collection of traits is a manifestation of the socialization process itself. In other words, these traits represent how we learn to act socially appropriately.

Implicit to Digman's (1997) work is the idea that what is and what is not socially desirable is multifaceted (see also Wood & Wortman, 2012). Furthermore, people may differ with regard to what behaviors or traits they find socially-desirable (de Vries et al., 2016). Thus, high levels of Conscientiousness may drive a person to be seen as hard-working and competent,

but high levels of Agreeableness may drive a person to be seen as considerate and modest. Although Conscientiousness and Agreeableness are related (or at least correlated), they are distinct constructs representing different behavioral proclivities, and it is reasonable to assert that high levels of one trait (i.e., Conscientiousness) may cause one to inflate self-ratings of some capacity or behavior while high levels of the other trait (i.e., Agreeableness) may cause one to deflate or at least constrain self-ratings of the same behavior or capacity. If the behaviors one is examining are relevant to both traits, it is reasonable to expect that divergent effects of those personality traits would be observed on self-ratings. Ultimately, we live in a paradox wherein it is socially desirable to be simultaneously braggadocios and humble. With Agreeableness, the socially-desirable tendency to be modest may prevail, restraining self-ratings of IL. Note too that Agreeableness was not among the Big 5 personality traits shown to be a significant predictor of self-ratings of leadership by Judge, LePine, and Rich (2006), nor was it found to be related to self-estimated ability by Visser et al. (2008). Indeed, Cheng et al. (2017) found agreeable people are more likely to make generous assessments of others. Agreeableness is also negatively related to narcissism (Graziano & Tobin, 2001). Because leaders low on Agreeableness (less modest) would be expected to be more likely to overrate, we expect SOA to be weaker at lower levels of Agreeableness.

H3 Leader Agreeableness will be more positively related to other-ratings than to self-ratings of IL.

H4 Self-other-ratings of IL will exhibit greater agreement at *higher* levels of Agreeableness than at *lower* levels of Agreeableness.

Openness to Experience and Self- and Other-Ratings of Leadership

Openness to Experience is a broad personality construct, characterized by both a particular psychic structure and motivation to seek out the unfamiliar (McCrae & Costa, 1997). Openness to Experience is seen in imaginativeness, aesthetic appreciation and sensitivity, depth of feeling, curiosity, creativity, and intellectuality (McCrae & Costa, 1989). Regarding the motivation to seek out the unfamiliar, Openness to Experience is seen in the proclivity to enlarge and examine experience and to be very contemplative about new experiences and ideas (McCrae & Costa, 1997). The FFM facets of Openness to Experience include Fantasy, Aesthetics, Feelings, Actions, Ideas, and Values (Costa et al., 1991).

Our rationale relating Openness to SOA is based on internality and observability. Fundamentally Openness is a matter of inner experience, a mental phenomenon related to the scope of awareness or the depth and intensity of consciousness” (McCrae & Costa, 1997, p. 835). Although there have been differences of opinion regarding the degree of observability of Openness, it was rated least observable among the Big 5 NEO PI-3 traits (de Vries et al., 2016). Connelly and Ones (2010) similarly characterize Openness as a trait low in visibility. Human and Biesanz (2011) found Openness to be lowest in observability by others among the Big 5 traits. Given its internality and low observability, we expect Openness to be more strongly related to self-ratings than to other ratings, with greater SOA at the lower end of the scale.

H5 Leader Openness to Experience will be more positively related to self-ratings than to other-ratings of IL.

H6 Self-other-ratings of IL will exhibit greater agreement at *lower* levels of Openness to Experience than at *higher* levels of Openness to Experience.

Extraversion and Self- and Other-Ratings of Leadership

As stated by Watson and Clark (1997, p. 767), “to most people, the term ‘Extravert’ quickly conjures up an image of one who seeks out and enjoys the companionship of others – one who is poised, confident, and facile in social situations.” Extraverted people seem especially capable of negotiating social hierarchies, emerging as leaders, and being effective as leaders (de Vries, 2012; Judge, Bono, Ilies, & Gerhardt, 2002). This personality trait is contrasted with the introvert, described as socially reserved, quiet, and thoughtful (Matthews, 2004). The FFM facets of Extraversion include Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, and Positive Emotions (Costa et al., 1991).

We expect Extraversion to be both internally and externally observable and thus related to both self- and other-ratings of leadership (Vazire, 2010). Regarding external observability, de Vries et al. (2016) assessed Extraversion to be highest among the Big 5 traits in observability. Despite this finding, we posit that the greater effect will be upon self-ratings due to self-enhancement. Supporting this notion, Visser et al. (2008) found Extraversion to be positively related to self-estimated ability. Similarly, Bell and Arthur (2008) found a positive relationship between participants’ degrees of Extraversion and their self-ratings of assessment center performance that was not found for other-ratings of their performance (i.e., the assessors working in the center). Regarding the Assertiveness aspect of Extraversion, Brutus et al. (1999) found dominance to be related to leaders’ self-ratings of their behavior and effectiveness, but not to ratings from subordinates, peers, or supervisors. These findings support some degree of self-enhancement as an artifact of Extraversion. However, extraverted people may also show a leniency bias toward others (Cheng et al., 2017). Because leaders may be more likely to overrate due to their Extraversion, we expect SOA to be greater at lower levels of Extraversion.

H7 Leader Extraversion will be more positively related to self-ratings than to other-ratings of IL.

H8 Self-other-ratings of IL will exhibit greater agreement at *lower* levels of Extraversion than at *higher* levels of Extraversion.

Neuroticism and Self- and Other-Ratings of Leadership

Neuroticism refers to the degree to which individuals experience psychological distress and are insecure, anxious, depressed, and emotional (Costa & McCrae, 1992a; Salgado, 2004). The opposite of Neuroticism is Emotional Stability, characterized as secure, non-anxious, calm, and self-confident. The FFM facets of Neuroticism include Anxiety, Hostility, Depression, Self-Consciousness, Impulsiveness, and Vulnerability (Costa et al., 1991).

Neuroticism may cause a divergence in SOA via several mechanisms. For example, de Vries et al. (2016) assessed Neuroticism (measured as Emotional Stability) as moderately observable, but stated that visible expressions of high Emotionality (i.e., Neuroticism) may be uncommon. This statement is supported by Digman's (1997) contention that Emotional Stability is a highly socially-desirable trait, which can be adapted to conclude that Neuroticism is highly *undesirable*. If outward expressions of Neuroticism are stifled, its effect on other-ratings may be muted.

Furthermore, Neuroticism is largely an internal state (Carlson et al., 2011; Funder & Dobroth, 1987; John & Robins, 1993; Norman & Goldberg, 1966). Because it is comprised of anxiety and self-doubt, we expect it to be negatively related to self-ratings of positive leadership behaviors embodied by IL. Leaders who doubt themselves, are anxious, and worry about how others view them are unlikely to rate themselves highly on positive leader behaviors. Indeed, Neuroticism is positively associated with a tendency to self-depreciate (Costa & McCrae,

1992a). Data by Judge et al. (2006) shows that Neuroticism is negatively related to self-ratings of leadership; also, Visser et al. (2008) found Neuroticism to be negatively related to self-estimated ability. Inversely, Emotional Stability has been shown to be related to generous self-assessments (Cheng et al., 2017). Therefore, as Neuroticism increases, self-ratings of positive leadership behaviors should decrease (less overrating) without as large a change in other-ratings, thereby increasing SOA.

H9 Leader Neuroticism will be more negatively related to self-ratings than to other-ratings of IL.

H10 Self-other-ratings of IL will exhibit greater agreement at *higher* levels of Neuroticism than at *lower* levels of Neuroticism.

Gender and Self- and Other-Ratings of Leadership

There is an abundance of literature that describes how and why men and women may differ on various facets of leadership (Antonakis, Avolio, & Sivasubramaniam, 2003; Eagly, Johannesen-Schmidt, & van Engen, 2003). For instance, because women may be more participative and considerate than men are, they may receive higher ratings from others at least on some leadership behaviors (Bass, Avolio, & Atwater, 1996). However, there is considerably less literature addressing SOA as it relates to gender compared to how gender relates to leadership in general (Fleenor et al., 2010). In their review and theorizing about SOA, Atwater and Yammarino (1997) hypothesized that SOA would be greater for females than males. Their support for this position came from a stream of research that focused on how men and women perceive, interpret, and use feedback about their behavior. For instance, Roberts and Nolen-Hoeksema (1989, p. 741) found that the women in their sample “seemed to respond in a way that indicated that they, more readily than the men, considered the external information, whether

positive or negative, to have self-evaluative meaning.” Roberts (1991, p. 297) stated that “women's self-evaluations are more responsive to the valence of the feedback they receive than are men's.” Thus, it appears that women perceive others’ evaluations to be more informational than do men, and are more likely to use this information when evaluating their own behavior (Roberts & Nolen-Hoeksema, 1994). Individuals who use such information in forming their self-perceptions should have self-ratings that are more congruent with others than those who do not.

Regarding leadership specifically, Moshavi, Brown, and Dodd (2003) found that male managers provided higher self-ratings of transformational leadership than did their female counterparts, despite no difference in follower ratings. Vecchio and Anderson (2009) similarly found that male managers more often provided overestimates of effectiveness than did females in a 360-degree feedback program. Beyond these findings, there is much research revealing a tendency for males to give higher self-estimates than females relative to others’ ratings (Jones & Fletcher, 2002; Lindeman, Sundvik, & Rouhiainen, 1995; Patiar & Mia, 2008; Visser et al., 2008). In addition, women appear to be more likely to receive higher other ratings than men as evidenced by the Paustian-Underdahl et al. (2014) meta-analysis, which could influence SOA.

Overall, we expect both males’ and females’ self-rating to be higher than others’ ratings due to the general tendency for self-enhancement. Women are less likely than men to inflate their self-ratings, and they tend to exhibit more congruence between self- and other-ratings via increased self-awareness relative to men (as demonstrated by Van Velsor et al. 1993). In other words, women are less likely to show self-other knowledge asymmetry and should show greater SOA than men.

H11 Being a female will be more positively related to other ratings than to self-ratings of IL.

H12 Self-other-ratings of IL will exhibit *greater* agreement for female leaders than for male leaders.

METHODS

Sample

The data were collected over six years from managers participating in company-sponsored training programs at seven multinational companies and two groups of managers attending an executive education course. The assessment was designed as an integral element to provide participants with useful diagnoses and self-insight for their leadership development.

HR offices identified bosses, peers, and subordinates as raters to give participants feedback. All data were collected anonymously via an online platform and only aggregated information was given to the managers to guarantee rater confidentiality. Participants completed self-assessments of their leadership behaviors and personality. Raters provided ratings on the leadership behaviors of focal managers. In total, data were collected from 448 managers (73.44% males; age = 38.60, SD = 6.54) and 3,315 raters, located in 30 countries (note, the listwise sample with full observations on all measures was 378 managers and 2,895 raters). The managers were employed in companies operating in the following sectors: banking (5.87%), insurance (39.36%), food manufacturing (25.18%), hospitality and retail (6.85%), telecommunications and high tech (15.65%), and the rest operating in other industries (7.09%).

Measures

Participants used a 5-point rating scale from 0 = *not at all* to 4 = *frequently if not always* (Bass, Cascio, & O'Connor, 1974) to rate the leader behaviors.

We gathered data on IL using the scales (8 items) of Antonakis and House (2014), and estimated a MIMIC model, where we regressed the factors on the fixed-effects of leaders, firms,

and countries (using the 12 cluster means of each of the scales, see Antonakis, Bendahan, Jacquart, & Lalive, 2010; Mundlak, 1978); we also controlled for the fixed effects of time (5 dummy variables), response language (2 dummy variables), as well as leader sex and age. Using MPlus, version 7.3 (Muthén & Muthén, 2012) we estimated the confirmatory model with a WLSMV estimator for modeling categorical (ordered) dependent variables and a cluster-robust estimate of the variance (at the leader level) on the “other” ratings of leadership.

The confirmatory factor analysis results for the available data not having missing values for independent variables (i.e., $N = 418$ leaders rated by $n = 3,164$ raters) provided support for the four factors, $\chi^2(98) = 114.59, p > .10$; for descriptive purposes we also report two indexes of fit, which are not tests of fit per se: RMSEA = .01 (90% CI from .00 to .01) and CFI = 1.00. Note, the higher order model failed to fit the data, $\chi^2(100) = 131.57, p < .05$, as indicated too by the difference test for the WLSMV estimator (Satorra & Bentler, 2001), $\chi^2(2) = 55.14, p < .001$, thus validating the decision to model four first-order factors as separate outcomes. Mean standardized loadings were .76. Given our regression modeling procedure (see below), we used the observed scale means as dependent variables because measurement errors are orthogonal to regressors and will not bias estimates (Ree & Carretta, 2006).

For personality, we used the 240 item NEO-PI-R self-personality assessment in English (Costa & McCrae, 1992b). The reliability of the five scales, using each of the respective six facets was satisfactory: Neuroticism $\alpha = .79$, Extraversion $\alpha = .76$, Openness to Experience $\alpha = .72$, Agreeableness $\alpha = .65$, Conscientiousness $\alpha = .76$. Note, because measurement errors in independent variables are not orthogonal to the regressors, we explicitly modeled measurement error (Ree & Carretta, 2006) using the reliabilities reported in the NEO-PI-R manual and global indexes of the Big Five factors (Costa & McCrae, 1992b).

Analysis method

Predicting (dis)agreement basically involves examining congruence as a dependent variable in the models estimated. Being cognizant of the drawbacks of using difference scores in studying congruence, we adopted the multivariate regression approach suggested by Edwards (1995) to test our hypotheses³. Although this approach was advanced almost two decades ago, it appears not to have been sufficiently understood and applied to SOA. We first generated an over-rater dummy variable (i.e., = 1 when one's self-rating is higher than other rating on the same construct, else = 0) for each dependent variable, and the interaction term of the over-rater dummy with the six independent variables of interest. We conducted multilevel analyses with 3,315 raters nested in 448 leader clusters (McNeish, Stapleton, & Silverman, 2017). In other words, the basic unit of analysis is each rater (n = 3,315), whereas the nested data structure is taken into account (leader cluster n = 448) in calculating the standard errors of the estimates. Specifically, we simultaneously estimated the following equations to predict self- and other rated leadership at the leader (*j*) and rater (*i*) level (note, for expositional clarity we do not subscript all regressors which are, of course, at the *j* level,):

$$\begin{aligned}
 y_{(S)j} = & a_0 + a_1C + a_2A + a_3O + a_4E + a_5N + a_6OVER + a_7OVER \times C \\
 & + a_8OVER \times A + a_9OVER \times O + a_{10}OVER \times E + a_{11}OVER \times N + a_{12}Age \\
 & + a_{13}GENDER + a_{14}\overline{C}_{country} + a_{15}\overline{A}_{country} + a_{16}\overline{O}_{country} + a_{17}\overline{E}_{country} \\
 & + a_{18}\overline{N}_{country} + a_{19}\overline{Age}_{country} + a_{20}\overline{Gender}_{country} + \sum_{m=2}^6 \gamma_{1m}Year_m \\
 & + \sum_{n=2}^3 \delta_{1n}Lang_n + \sum_{k=2}^9 \omega_{1k}Firm_k + \varepsilon_j
 \end{aligned}$$

³ Whereas polynomial regression is widely adopted as an adequate method for studying SOA as predictor (Edwards, 1994; Edwards & Parry, 1993), we used multivariate regression in this study which treats SOA as a dependent variable (Edwards, 1995).

$$\begin{aligned}
y_{(o)ij} = & \beta_0 + \beta_1 C + \beta_2 A + \beta_3 O + \beta_4 E + \beta_5 N + \beta_6 OVER + \beta_7 OVER \times C \\
& + \beta_8 OVER \times A + \beta_9 OVER \times O + \beta_{10} OVER \times E + \beta_{11} OVER \times N + \beta_{12} Age \\
& + \beta_{13} GENDER + \beta_{14} \overline{C}_{country} + \beta_{15} \overline{A}_{country} + \beta_{16} \overline{O}_{country} + \beta_{17} \overline{E}_{country} \\
& + \beta_{18} \overline{N}_{country} + \beta_{19} \overline{Age}_{country} + \beta_{20} \overline{Gender}_{country} + \sum_{m=2}^6 \gamma_{2m} Year_m \\
& + \sum_{n=2}^3 \delta_{2n} Lang_n + \sum_{k=2}^9 \omega_{2k} Firm_k + \mu_{ij}
\end{aligned}$$

Where $y_{(s)}$ = self-rated sub-dimension of IL, $y_{(o)}$ = other-rated sub-dimension of IL, C = Conscientiousness, A = Agreeableness, O = Openness to Experience, E = Extraversion, N = Neuroticism, $OVER$ = over-rater dummy, Age = age, $Gender$ = gender dummy, $Year$ = dummies for the year of data collection, $Lang$ = language dummies, $Firm$ = organization dummies. Note that we included country-level cluster means for all regressors, indicated by a “bar” for C , A , O , E , N , Age , and $Gender$ so as to control for the unobserved heterogeneity at the country level, (Antonakis et al., 2010; Mundlak, 1978); in this way, we avoided introducing a huge number of country dummies as controls. To improve estimation efficiency, we allowed the disturbances (ε and μ) of the two equations to correlate.

We estimated the model using Stata’s (StataCorp, 2015) maximum likelihood estimator for missing data and thus maximized the available sample size; we estimated the models using a cluster-robust estimator (Rogers, 1994) to ensure correct standard errors at the leader level and to be able to estimate and make meaningful cross-equation tests for models having dependent variables at the ij and j level simultaneously. We modelled the personality variables as latent variables by using the scale as an indicator of a latent variable and constraining its disturbance to the formula provided by Bollen (1989); that is, for the latent variable ξ the disturbance of the indicator is constrained to $(1 - \rho) / \text{var}_x$, where ρ is an estimate of the reliability and var_x is the

observed variance of the indicator. In so doing, we addressed the issue of measurement error in the independent variables.

RESULTS

For computational convenience in estimation, we reduced the variance of the variables to ensure convergence of the estimators by rescaling the scores of the five personality dimensions (i.e., by dividing them by a constant; specifically, the scores are divided by 60, 30, 60, 70, and 30 for Conscientiousness, Agreeableness, Openness, Extraversion, and Neuroticism, respectively, which brought their means close to unity). The raw means, standard deviations, and correlations are presented in Table 1. We first examined whether self-ratings are indeed higher than other-ratings. Cluster-robust tests using SEM showed that the mean self-ratings across the scales (2.97) are +13.36% higher than the mean other-ratings (2.62), $\chi^2(1) = 183.45$, $p < .001$. This result is consistent with our expectations and allows us to make inferences about agreement as discussed in our hypotheses.

[Insert Table 1 about here]

Because self-ratings may be higher or lower than other-ratings (resulting in over- or under-raters respectively), we first examined the effect of the personality dimension on leadership for *over-raters* versus *under-raters* (Edwards, 1995). The results of multivariate (omnibus) tests, as well as tests of each individual interaction term (using Wald test with Bonferroni-adjusted p-values), showed the interaction terms to be non-significant for all equations. Thus, we concluded that it was appropriate to drop the over-rater dummy and interaction terms (i.e., the variables with coefficients $\alpha_6 - \alpha_{11}$ and $\beta_6 - \beta_{11}$ in the equations above).

Subsequently, we re-ran the multivariate regressions using SEM. The regression results are reported in Table 2. To test our odd-numbered hypotheses, which suggested that the effect of leaders' personality traits on self- and other-ratings of leadership behaviors will differ, we compared whether the coefficients of the same personality traits were significantly different in predicting self- versus other-rated leadership.

To test the even-numbered hypotheses that explicitly spelled out *when* (i.e., at higher versus lower levels of specific personality traits) greater agreement will happen. For each personality factor and gender we simultaneously compared at low and high levels of each factor, whether the point estimate of the other rating is different from the 95% lower-bound value of the self-rating. Thus, we could determine whether the difference in predictions for self-and other-ratings differed significantly at the low and high level of each factor. The results are reported in Table 3.

[Insert Table 2 & 3 about here]

Our Hypotheses 1 and 2 stated that leader Conscientiousness will be more positively related to self-ratings of IL compared to other-ratings (H1) and consequently the agreement will be greater at *lower* levels of Conscientiousness (H2). We found that Conscientiousness more strongly predicts self-ratings of IL ($\beta = .27, .40, .41, .28$ for self-rated EM, SF, PG, and OM respectively, all $p < .001$) than other-ratings ($\beta = -.00, .12, .14, .05$ for other-rated EM, SF, PG, and OM respectively, p ranges from .05 to n.s.). The result of the Wald test showed that all these four pairs of coefficients are significantly different ($\chi^2(1) = 5.61, 6.45, 5.16, 4.80$, all $p < .05$, for EM, SF, PG, and OM respectively). Furthermore, there was greater agreement at low than at high levels of Conscientiousness (see Table 3) for all the four sub-dimensions of IL. Figure 1

illustrates such a relationship with regard to the SF sub-dimension of IL. Taken together, these results supported our Hypotheses 1 and 2.

[Insert Figure 1 about here]

In Hypotheses 3 and 4, we expected that leader Agreeableness would be more positively related to other-ratings of IL compared to self-ratings (H3) and consequently the agreement will be greater at *higher* levels of Agreeableness (H4). The results showed that Agreeableness is significantly related to other-ratings of IL ($\beta = .06, .09, .07, .08, p < .05, .01, .10, .05$ for other-rated EM, SF, PG, and OM respectively) yet not significantly to self-ratings ($\beta = -.04, .02, .06, .10$, for self-rated EM, SF, PG, and OM respectively; all n.s. except that of OM $p < .10$); however, we only found a marginal difference between these two coefficients for EM ($\chi^2(1) = 3.02, p < .10$), not between other pairs. Also, there was greater agreement at higher levels than at lower levels of Agreeableness for EM and SF (see Table 3). Figure 2 shows the relationship between Agreeableness and SOA. Hence Hypotheses 3 and 4 are only partially supported.

[Insert Figure 2 about here]

We expected that leader Openness to Experience would be more positively related to self-ratings of IL as compared to other-ratings (H5) and consequently, the agreement will be greater at *lower* levels of Openness (H6). The relationship between Openness and self-ratings ($\beta = .13, .07, .25, .09$, for self-rated EM, SF, PG, and OM respectively, were all n.s. except that of PG $p < .05$) and were not consistently stronger than that between Openness and other-ratings ($\beta = .12, .08, .08, .11$, for other-rated EM, SF, PG, and OM respectively, all n.s. except that of EM $p < .05$). In addition, they are not significantly different from each other (Wald test $\chi^2(1)$ all n.s.). The test of magnitude difference indicates that the pattern of greater agreement at higher levels of Openness only exists for PG (see Table 3). As a result, Hypotheses 5 and 6 are not supported.

Hypotheses 7 and 8 stated that leader Extraversion would be more positively related to self-ratings of IL compared to other-ratings (H7) and consequently the agreement will be greater at *lower* levels of Extraversion (H8). We found that Extraversion is neither related to self-ratings ($\beta = .11, .04, .05, .00$, for self-rated EM, SF, PG, and OM respectively, all n.s.) nor other ratings ($\beta = .08, .01, .08, -.02$, for other-rated EM, SF, PG, and OM respectively, all n.s.), and that they do not significantly differ from each other (all Wald test n.s.). There was no difference in agreement at low or high levels of Extraversion (see Table 3). Hence Hypotheses 7 and 8 are not supported.

Hypotheses 9 and 10 proposed that leader Neuroticism will be more highly *negatively* related to self-ratings of IL compared to other-ratings (H9) and consequently the agreement will be greater at *higher* levels of Neuroticism (H10). We found a negative coefficient for Neuroticism predicting self-ratings of IL ($\beta = -.05, -.08, -.06, -.04$, for self-rated EM, SF, PG, and OM respectively), though not significant. The relationship for Neuroticism and other-ratings is also almost zero (though generally positive, $\beta = .02, .06, .05, -.01$, for other-rated EM, SF, PG, and OM respectively, all n.s., except for PG $p < .10$). Interestingly, the Wald test suggested that these two coefficients differ from each other for the SF and PG sub-dimensions ($\chi^2(1) = 5.75, p < .05$; $\chi^2(1) = 2.75, p < .10$ respectively). In other words, although the two coefficients are not significantly different from zero individually, the difference between them is statistically significant (Gelman & Stern, 2006). Also, agreement was greater at higher levels of Neuroticism (for EM, SF, and PG, see Table 3). Figure 3 illustrates the relationship between Neuroticism and SOA. Hence, Hypotheses 9 was not supported yet Hypothesis 10 was generally supported.

[Insert Figure 3 about here]

Hypotheses 11 and 12 proposed that a leader being female versus male will be more positively related to other-ratings of IL compared to self-ratings (H11) and consequently the agreement will be greater for females versus males (H12). We found that other ratings of IL ($\beta = .03, .10, .13, .12$, for other-rated EM, SF, PG, and OM respectively, n.s., $p < .10, .01, .05$) were more strongly related to being a female than were self-ratings ($\beta = -.19, -.01, .08, .03$, for self-rated EM, SF, PG, and OM respectively, all n.s. except EM $p < .01$); we only found a difference between the two coefficients for EM ($\chi^2(1) = 5.85, p < .05$). Also, agreement coincided with being female (coded 1) rather than being male (coded 0) in all sub-dimensions except for PG, as indicated in Table 3. Figure 4 shows the relationship between gender and SOA. These results generally support Hypotheses 11 and 12.

Interested readers may refer to the supplementary materials, where we also report results for transformational leadership, using the Multifactor Leadership Questionnaire.

[Insert Figure 4 about here]

DISCUSSION

Findings and Contributions

Our paper is among the first to address SOA from the perspective of trait-induced divergence, focusing on some of the most widely studied individual differences, the Big Five personality traits and gender. Furthermore, we applied advanced statistical methods to predict congruence (e.g., Edwards, 1995), which allowed us to investigate predictors of agreement in a more rigorous and multivariate way. Additionally, we used robust controls, corrected for the effects of measurement error, and modeled fixed-effects; given that our independent variables are largely exogenous, this suggests that our estimates are largely unconfounded.

Our results imply that we may need to reconsider how SOA is conceptualized and whether certain personality traits always contribute positively to leadership outcomes. Take leader Conscientiousness, for instance, which decreased SOA, and leader Neuroticism, which increased SOA. Although Conscientiousness is heralded as highly desirable (McFarland & Ryan, 2000; Stöber, 2001) and predictive of various important leadership and performance outcomes (Barrick et al., 2001; Hertz & Donovan, 2000), there is at least one crucial aspect associated with increasing levels of the trait for leaders we would like to entertain; “whereas conscientious individuals may be diligent in their work and attentive to detail, highly conscientious leaders may emerge as perfectionists, inflexible about procedures and policies, and critical of their team's performance” (Judge, Piccolo, & Kosalka, 2009, p. 868). Thus, such leaders’ self-insight may be clouded by this “dark side” of Conscientiousness (Hogan & Hogan, 2001; Pierce & Aguinis, 2011; Tett, 1998). This peculiarity is likely due at least in part, to an increased tendency for highly conscientious individuals to engage in self-enhancement as a result of its high social desirability and a desire to maintain a positive self-image (Lee & Klein, 2002; Martocchio & Judge, 1997; Stöber et al., 2002). This tendency to rate oneself highly may explain why higher Conscientiousness is associated with decreased SOA (Fleenor et al., 2010). It seems that high Conscientiousness may be a mixed blessing. Although it contributes to improved job performance, it may in some circumstances detract from effective leader-follower relationships. Conversely, although Neuroticism is generally viewed unfavorably, there seems to be at least some utility associated with higher levels of the trait (Judge & LePine, 2007) in terms of greater SOA. Moreover, “individuals low on emotional stability, because they worry about meeting expectations, may actually *exceed* them” (Judge & LePine, 2007, p. 337). This preoccupation with monitoring the environment, including other’s expectations, may allow those higher on

neuroticism greater insight into how others see them, thereby resulting in increased self-awareness and SOA as Neuroticism increases. Somewhat surprisingly, our results showed that there was no effect of Neuroticism on other's ratings, suggesting that Neuroticism may, at least in certain situations, manifest itself largely internally.

Our hypotheses for gender were also supported in that being female was more strongly related to other ratings than to self-ratings resulting in greater agreement for females as self-ratings were less inflated relative to others'.

Theoretical Implications

With the exception of de Vries et al. (2016), explorations of the effect of personality on self- and other-ratings are conspicuously lacking in the literature. We build upon a handful of studies (see Fleenor et al., 2010) connecting Big 5 personality factors to SOA research, exploring how a leader's personality may affect self- and other-ratings of their IL behaviors. We advance the knowledge on SOA in the following ways. First, our study showcased the importance of applying adequate methods to study the prediction of SOA (e.g., Edwards, 1995) in that such methods allow us to depict a more refined understanding of SOA that has not yet received sufficient attention. Whereas conventional studies associate individual differences with SOA directly, our results suggest that leader personality and gender may cast a differential influence on self- and other-ratings of leadership behavior, which in turn causes greater or lesser SOA. Even though some personality traits do not seem to be directly associated with self- or other-ratings (e.g., Conscientiousness is not significantly related to other-ratings of IL), this results does not imply that such traits do not predict SOA. Thus, theories regarding SOA should be revisited to consider these complex multivariate relations and question the value of SOA per se concerning how it has been traditionally modeled.

Second, our results support the notion that various aspects of personality and gender differentially affect self- and other-ratings of IL, which produce varying degrees of SOA. This idea provides a deeper understanding of how individual differences affect SOA in leadership assessment from the perspective of information asymmetry (Vazire, 2010), as well as self-enhancement and social desirability biases. Thus, we hope that current theory will be refined to consider these perspectives and the relationship of information asymmetry to self-and other-ratings. Additionally, theory (and empirical studies) may want to revisit how personality traits may be predictive in other ways (for SOA), which may not follow typical main effects findings studied in isolation.

Practical Implications

SOA is a commonly used measure to operationalize self-awareness, and a great deal of research has suggested that SOA is related to positive outcomes. However, it is interesting to consider from this study that agreement may stem from factors other than the self and others merely viewing the leader's behavior similarly. For instance, in our data, agreement on IL is greater when leader Conscientiousness is low and leader Neuroticism is high, yet these leader personality characteristics do not seem to influence follower ratings on IL. In light of these findings, perhaps we should reflect on the suggestion that agreement is a good indicator of self-awareness. It merely may be, at least in part, an artifact of leader personality, gender, or other characteristics of the leader. Also, though not examined in this study, we could speculate that additional self (or other) rater characteristics also may influence SOA.

Limitations and Future Research

Despite the strengths of our study, our findings have notable limitations. First, we collected data in a cross-sectional fashion. Nevertheless, given that personality traits and gender

tend to be stable and can be considered exogenous, we feel optimistic about the predictive effect of these individual differences because the personality-leader relationship can be assumed to be in equilibrium over time (Fischer, Dietz, & Antonakis, 2016). Yet, we encourage future research to study such phenomena with designs that allow for stronger causal claims. Second, we focused only on leaders' personality and gender. However, followers' personality and gender are also likely to influence how they process information (hence affect their perception of leaders), which in turn affects SOA. Relatedly, the gender match between leaders and subordinates may influence how they rate the leaders. Unfortunately, for the sake of confidentiality, we were not able to collect any personal information about subordinates who rated the leaders. One thing to note, though, is that most of the leaders were males, and given the industries we sampled, we can safely state too that most of the respondents were males. Still, the fact we could not control for team gender composition is a limitation. Thus, future research should consider ways to incorporate followers' individual differences in predictive models of the sort we have used. Third, although we have collected data from multiple sources, leader personality traits and self-ratings do come from the same source (i.e., the leader in question). As a result, there may be a risk of common method variance in our analyses. However, because personality and gender are largely exogenous, and because we controlled for all fixed-effects at the country, firm, and time level, this bias should be minimal. In any case, there were many more instances of significant associations for the leader individual difference with other ratings than with self-ratings, suggesting that same source effects were trivial. Also, self-rated personality scales are not highly prone to desirability bias given the high self-other concordance in ratings that are observed (Connolly, Kavanagh, & Viswesvaran, 2007b).

It should also be noted that because data were gathered anonymously, we were not able to isolate the rater group (e.g., boss, peer, subordinate) from which the ratings were obtained, though most of the ratings came from the leader's subordinates. Future research may consider mitigating such risk by assessing leader personality using other-ratings from one subset of raters (and another subset for leadership behaviors). Fourth, an idea for future research suggested by an anonymous reviewer is also provocative; that is, it would be worthwhile to assess self-and other-ratings of both personality and leadership to disentangle the relationships better. Finally, we used a leadership scale that is not emotion-laden but focused on technical aspects of leading. Our results, therefore, may be bounded by the nature of the scale we used. Future research may wish to examine how our modeling procedure would explain SOA when using scales capturing socio-emotional aspects of leader behavior (see supplementary materials).

CONCLUSIONS

Our results suggest that scholars studying SOA may want to reconsider what it means to study this phenomenon. There seem to be heretofore unexpected and understudied factors such as personality and gender that influence self-and other-ratings, which contribute to agreement or disagreement. Additionally, personality traits that have long been considered "good" and "bad" may actually have mixed results in organizational settings because even though they contribute positively to some outcomes, they may contribute negatively to others. We hope that scholars test some of the ideas we have proposed in large-scale studies to determine whether our findings can be replicated and further explained.

REFERENCES

- Antonakis, J., Avolio, B. J., & Sivasubramaniam, N. (2003). Context and leadership: an examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly*, *14*(3), 261–295.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120.
- Antonakis, J., & House, R. J. (2002). The full-range leadership theory: The way forward. In B. J. Avolio & F. J. Yammarino (Eds.), *Transformational and charismatic leadership: The road ahead* (Vol. 2, pp. 3–34). Amsterdam: JAI Press.
- Antonakis, J., & House, R. J. (2014). Instrumental leadership: Measurement and extension of transformational–transactional leadership theory. *The Leadership Quarterly*, *25*(4), 746–771. <https://doi.org/10.1016/j.leaqua.2014.04.005>
- Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360 feedback. *Journal of Vocational Behavior*, *66*(3), 532–548.
- Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, *45*(1), 141–164.
- Atwater, L. E., & Yammarino, F. J. (1997). Self-other rating agreement: A review and model. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 15, pp. 121–174). JAI Press.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next? *International Journal of Selection and Assessment*, *9*(1&2), 9–30. <https://doi.org/10.1111/1468-2389.00160>

- Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology, 53*(4), 925–951.
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: The Free Press.
- Bass, B. M., Avolio, B. J., & Atwater, L. (1996). The Transformational and Transactional Leadership of Men and Women. *Applied Psychology, 45*(1), 5–34. <https://doi.org/10.1111/j.1464-0597.1996.tb00847.x>
- Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology, 59*(3), 313–320.
- Bass, B. M., & Yammarino, F. J. (1991). Congruence of Self and Others' Leadership Ratings of Naval Officers for Understanding Successful Performance. *Applied Psychology, 40*(4), 437–454. <https://doi.org/10.1111/j.1464-0597.1991.tb01002.x>
- Bell, S. T., & Arthur, W. (2008). Feedback acceptance in developmental assessment centers: the role of feedback message, participant personality, and affective response to the feedback session. *Journal of Organizational Behavior, 29*(5), 681–703. <https://doi.org/10.1002/job.525>
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bono, J. E., & Judge, T. A. (2004). Personality and Transformational and Transactional Leadership: A Meta-Analysis. *Journal of Applied Psychology, 89*(5), 901–910.
- Brutus, S., Fleenor, J. W., & McCauley, C. D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development, 18*(5), 417–435. <https://doi.org/10.1108/02621719910273569>

- Burris, C. T., & Navara, G. S. (2002). Morality Play or Playing Morality?: Intrinsic Religious Orientation and Socially Desirable Responding. *Self and Identity*, 1(1), 67–76. <https://doi.org/10.1080/152988602317232812>
- Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impressions: Is negative affectivity associated with greater realism? *Journal of Personality and Social Psychology*, 58(1), 122–133.
- Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology*, 101(4), 831–846.
- Cavazotte, F., Moreno, V., & Hickmann, M. (2012). Effects of leader intelligence, personality and emotional intelligence on transformational leadership and managerial performance. *The Leadership Quarterly*, 23(3), 443–455. <https://doi.org/10.1016/j.leaqua.2011.10.003>
- Cheng, K. H. C., Hui, C. H., & Cascio, W. F. (2017). Leniency Bias in Performance Ratings: The Big-Five Correlates. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00521>
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136(6), 1092–1122. <https://doi.org/10.1037/a0021212>
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007a). The Convergent Validity between Self and Observer Ratings of Personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15(1), 110–117. <https://doi.org/10.1111/j.1468-2389.2007.00371.x>
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007b). The Convergent Validity between Self and Observer Ratings of Personality: A meta-analytic review. *International Journal*

- of Selection and Assessment*, 15(1), 110–117. <https://doi.org/10.1111/j.1468-2389.2007.00371.x>
- Costa, P. T., & McCrae, R. R. (1985). *The NEO-PI manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1992a). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4(1), 5–13. <https://doi.org/10.1037/1040-3590.4.1.5>
- Costa, P. T., & McCrae, R. R. (1992b). *The NEO-PI professional manual*. Lutz, FL.: Psychological Assessment Resources.
- Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory. *Personality and Individual Differences*, 12(9), 887–898. [https://doi.org/10.1016/0191-8869\(91\)90177-D](https://doi.org/10.1016/0191-8869(91)90177-D)
- de Vries, R. E. (2012). Personality predictors of leadership styles and the self–other agreement problem. *The Leadership Quarterly*, 23(5), 809–821.
- de Vries, R. E., Realo, A., & Allik, J. (2016). Using Personality Item Characteristics to Predict Single-Item Internal Reliability, Retest Reliability, and Self-Other Agreement: Item characteristics, reliability, and self-other agreement. *European Journal of Personality*, 30(6), 618–636. <https://doi.org/10.1002/per.2083>
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73, 1246–1256.
- Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of*

- Personality and Social Psychology*, 57(6), 1082–1090. <https://doi.org/10.1037/0022-3514.57.6.1082>
- Eagly, A. H., Johannesen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, 129(4), 569–591.
- Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, 58, 51–100.
- Edwards, J. R. (1995). Alternatives to Difference Scores as Dependent Variables in the Study of Congruence in Organizational Research. *Organizational Behavior and Human Decision Processes*, 64(3), 307–324.
- Edwards, J. R., & Parry, M. E. (1993). On the Use of Polynomial Regression Equations as an Alternative to Difference Scores in Organizational Research. *The Academy of Management Journal*, 36(6), 1577–1613.
- Emmons, R. A. (1987). Narcissism: theory and measurement. *Journal of Personality and Social Psychology*, 52(1), 11.
- Fischer, T., Dietz, J., & Antonakis, J. (2016). Leadership Process Models: A Review and Synthesis. *Journal of Management*, 43(6), 1726–1753. <https://doi.org/10.1177/0149206316682830>
- Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self–other rating agreement in leadership: A review. *The Leadership Quarterly*, 21(6), 1005–1034.
- Funder, D. C. (1980). On seeing ourselves as others see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality*, 48(4), 473–493.

- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177–182. <https://doi.org/10.1177/0963721412445309>
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55(1), 149–158. <https://doi.org/10.1037/0022-3514.55.1.149>
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52(2), 409–418.
- Furnham, A., & Stringfield, P. (1998). Congruence in job-performance ratings: a study of 360 degrees feedback examining self, manager, peers, and consultant ratings. *Human Relations*, 51, 517–530.
- Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*, 60(4), 328–331. <https://doi.org/10.1198/000313006X152649>
- Goffin, R. D., & Anderson, D. W. (2007). The self-rater’s personality and self-other disagreement in multi-source performance ratings: Is disagreement healthy? *Journal of Managerial Psychology*, 22(3), 271–289. <https://doi.org/10.1108/02683940710733098>
- Graziano, W. G., & Tobin, R. M. (2001). The distributed processing of narcissists: Paradox lost? *Psychological Inquiry*, 12, 219–222.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41(1), 43–62. <https://doi.org/10.1111/j.1744-6570.1988.tb00631.x>
- Hogan, R., & Hogan, J. (2001). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment*, 9(1/2), 40–51.

- Human, L. J., & Biesanz, J. C. (2011). Target adjustment and self-other agreement: Utilizing trait observability to disentangle judgeability and self-knowledge. *Journal of Personality and Social Psychology, 101*(1), 202–216. <https://doi.org/10.1037/a0023782>
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: the Big Five revisited. *Journal of Applied Psychology, 85*(6), 869–879.
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*(4), 521–551.
- John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*(1), 206–219. <https://doi.org/10.1037/0022-3514.66.1.206>
- Jones, L., & Fletcher, C. (2002). Self-assessment in a selection situation: An evaluation of different measurement approaches. *Journal of Occupational and Organizational Psychology, 75*(2), 145–161.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: a qualitative and quantitative review. *Journal of Applied Psychology, 87*(4), 765.
- Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (2006). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621–652.
- Judge, T. A., & LePine, J. A. (2007). The bright and dark sides of personality: implications for personnel selection in individual and team contexts. In C. C. Langan-Fox & R. J. Klimoski (Eds.), *Research companion to the dysfunctional workplace: Management challenges and symptoms* (pp. 332–355). Cheltenham, UK: Edward Elgar.

- Judge, T. A., LePine, J. A., & Rich, B. L. (2006). Loving Yourself Abundantly: Relationship of the Narcissistic Personality to Self- and Other Perceptions of Workplace Deviance, Leadership, and Task and Contextual Performance. *Journal of Applied Psychology, 91*(4), 762–776. <https://doi.org/10.1037/0021-9010.91.4.762>
- Judge, T. A., Piccolo, R. F., & Kosalka, T. (2009). The bright and dark sides of leader traits: A review and theoretical extension of the leader trait paradigm. *The Leadership Quarterly, 20*, 855–875. <https://doi.org/10.1016/j.leaqua.2009.09.004>
- Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing Individual Differences in Self-Enhancement Bias: An Interpersonal Approach. *Psychological Review, 111*(1), 94–110. <https://doi.org/10.1037/0033-295X.111.1.94>
- Kwan, V. S. Y., John, O. P., Robins, R. W., & Kuang, L. L. (2008). Conceptualizing and assessing self-enhancement bias: A componential approach. *Journal of Personality and Social Psychology, 94*(6), 1062–1077. <https://doi.org/10.1037/0022-3514.94.6.1062>
- Lee, S., & Klein, H. J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology, 87*(6), 1175–1182.
- Lindeman, M., Sundvik, L., & Rouhiainen, P. (1995). Under- or overestimation of self? Person variables and self-assessment accuracy in work settings. *Journal of Social Behavior & Personality, 10*(1), 123–134.
- Lord, R. G., de Vader, C. L., & Alliger, G. M. (1986). A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology, 71*(3), 402–410.

- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280–296. <https://doi.org/10.1037/0021-9010.67.3.280>
- Malik, K. B., Hussain, L., Ali, U., & Ali, M. (2011). Principal's personal traits as determinants of leadership effectiveness. *Interdisciplinary Journal of Contemporary Research In Business, 3*(2), 1565–1569.
- Martocchio, J. J., & Judge, T. A. (1997). Relationship between conscientiousness and learning in employee training: mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology, 82*(5), 764–773.
- Matthews, G. (2004). Extroversion-Introversion. In C. D. Spielberger (Ed.), *Encyclopedia of Applied Psychology* (Vol. 1, pp. 869–873). Oxford: Academic Press.
- McCall, M. W., & Lombardo, M. M. (1983). *Off the track: Why and how successful executives get derailed*. Center for Creative Leadership.
- McCrae, R. R., Costa, Jr., P. T., & Martin, T. A. (2005). The NEO-PI-3: A More Readable Revised NEO Personality Inventory. *Journal of Personality Assessment, 84*(3), 261–270. https://doi.org/10.1207/s15327752jpa8403_05
- McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality. *Journal of Personality, 57*(1), 17–40.
- McCrae, R. R., & Costa, P. T. (1997). Conceptions and correlates of openness to experience. In R. Hogan, J. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 825–847). San Diego: Academic Press.

- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*(5), 812–821.
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114–140. <https://doi.org/10.1037/met0000078>
- Moshavi, D., Brown, F. W., & Dodd, N. G. (2003). Leader self-awareness and its relationship to subordinate attitudes and performance. *Leadership & Organization Development Journal, 24*(7), 407–418. <https://doi.org/10.1108/01437730310498622>
- Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resources Management, 13*(3), 153–200.
- Mount, M. K., Ilies, R., & Johnson, E. (2006). Relationship of personality traits and counterproductive work behaviors: The mediating effects of job satisfaction. *Personnel Psychology, 59*(3), 591–622.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society, 1*(46), 69–85.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide (7th ed.)*. Los Angeles, CA: Muthén & Muthén.
- Noles, S. W., Cash, T. F., & Winstead, B. A. (1985). Body image, physical attractiveness, and depression. *Journal of Consulting and Clinical Psychology, 53*(1), 88–94.
- Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4*(6), 681–691.

- Ogunfowora, B., Bourdage, J., & Lee, K. (2010). Rater personality and performance dimension weighting in making overall performance judgments. *Journal of Business and Psychology, 25*(3), 465–476.
- Patiar, A., & Mia, L. (2008). The effect of subordinates' gender on the difference between self-ratings, and superiors' ratings, of subordinates' performance in hotels. *International Journal of Hospitality Management, 27*(1), 53–64. <https://doi.org/10.1016/j.ijhm.2007.07.009>
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*(2), 307.
- Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology, 99*(6), 1129–1145. <https://doi.org/10.1037/a0036751>
- Pierce, J. R., & Aguinis, H. (2011). The Too-Much-of-a-Good-Thing Effect in Management. *Journal of Management, 39*(2), 313–338. <https://doi.org/10.1177/0149206311410060>
- Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management, 12*(4), 531–544. <https://doi.org/10.1177/014920638601200408>
- Pronin, E. (2008). How we see ourselves and how we see others. *Science (New York, N.Y.), 320*(5880), 1177–80. <https://doi.org/10.1126/science.1154199>
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The Bias Blind Spot: Perceptions of Bias in Self Versus Others. *Personality and Social Psychology Bulletin, 28*(3), 369–381. <https://doi.org/10.1177/0146167202286008>

- Rauthmann, J. F. (2012). You Say the Party is Dull, I Say It is Lively: A Componential Approach to How Situations Are Perceived to Disentangle Perceiver, Situation, and Perceiver \times Situation Variance. *Social Psychological and Personality Science*, 3(5), 519–528. <https://doi.org/10.1177/1948550611427609>
- Ree, M. J., & Carretta, T. R. (2006). The role of measurement error in familiar statistics. *Organizational Research Methods*, 9(1), 99–112.
- Roberts, T. A. (1991). Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin*, 109(2), 297–308. <https://doi.org/10.1037/0033-2909.109.2.297>
- Roberts, T. A., & Nolen-Hoeksema, S. (1989). Sex differences in reactions to evaluative feedback. *Sex Roles*, 21(11), 725–747.
- Roberts, T. A., & Nolen-Hoeksema, S. (1994). Gender comparisons in responsiveness to others' evaluations in achievement settings. *Psychology of Women Quarterly*, 18(2), 221–240.
- Rogers, W. (1994). Regression standard errors in clustered samples. *Stata Technical Bulletin*, 3(13), 19–23.
- Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76(3), 323–346. <https://doi.org/10.1348/096317903769647201>
- Salgado, J. F. (2004). Traits. In C. D. Spielberger (Ed.), *Encyclopedia of Applied Psychology* (Vol. 3, pp. 569–573). Oxford: Academic Press.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. <https://doi.org/10.1007/BF02296192>

- Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the Big Five. *Journal of Research in Personality, 38*(5), 473–480.
<https://doi.org/10.1016/j.jrp.2003.09.010>
- Sedikides, C., & Gregg, A. P. (2008). Self-Enhancement: Food for Thought. *Perspectives on Psychological Science, 3*(2), 102–116. <https://doi.org/10.1111/j.1745-6916.2008.00068.x>
- Sinha, N., Mesmer-Magnus, J., & Viswesvaran, C. (2012). Personality Antecedents of Self-Other Rating Discrepancy. *Revista de Psicología Del Trabajo y de Las Organizaciones, 28*(1), 3–3.
- Smither, J. W., Brett, J. F., & Atwater, L. E. (2007). What Do Leaders Recall About Their Multisource Feedback? *Journal of Leadership & Organizational Studies, 14*(3), 202–218.
<https://doi.org/10.1177/1071791907308463>
- StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX.: StataCorp LP.
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*(3), 222–232.
- Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment, 78*(2), 370–389.
- Taylor, S. E. (1989). *Positive illusions: Creative self-deception and the healthy mind*. New York: Basic Books.
- Tett, R. P. (1998). Is conscientiousness always positively related to job performance. *The Industrial-Organizational Psychologist, 36*(1), 24–29.

- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88(3), 500–517. <https://doi.org/10.1037/0021-9010.88.3.500>
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel Psychology*, 44(4), 703–742.
- Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick, and Hunter (1994). *Personnel Psychology*, 47(1), 157–172.
- Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology*, 33(2), 263–271. <https://doi.org/10.1111/j.1744-6570.1980.tb02348.x>
- Van Velsor, E., Taylor, S., & Leslie, J. B. (1993). An examination of the relationships among self-perception accuracy, self-awareness, gender, and leader effectiveness. *Human Resource Management*, 32(2–3), 249–263. <https://doi.org/10.1002/hrm.3930320205>
- Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281–300. <https://doi.org/10.1037/a0017908>
- Vecchio, R. P., & Anderson, R. J. (2009). Agreement in Self-Other Ratings of Leader Effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment*, 17(2), 165–179. <https://doi.org/10.1111/j.1468-2389.2009.00460.x>

- Vernon, P. A., Villani, V. C., Vickers, L. C., & Harris, J. A. (2008). A behavioral genetic investigation of the Dark Triad and the Big 5. *Personality and Individual Differences, 44*(2), 445–452.
- Visser, B. A., Ashton, M. C., & Vernon, P. A. (2008). What makes you think you're so smart? Measured abilities, personality, and sex differences in relation to self-estimates of multiple intelligences. *Journal of Individual Differences, 29*(1), 35–44.
- Watson, D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In R. Hogan, J. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 767–793). San Diego: Academic Press.
- Wood, D., & Wortman, J. (2012). Trait Means and Desirabilities as Artifactual and Real Sources of Differential Stability of Personality Traits: Trait Desirability and Stability. *Journal of Personality, 80*(3), 665–701. <https://doi.org/10.1111/j.1467-6494.2011.00740.x>

Table 1
Means , Standard Deviations, and Correlations

Var.	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. C	60.41	25.74														
2. A	38.96	25.84	.09													
3. O	63.81	25.90	-.06	-.05												
4. E	72.66	23.18	.21	.11	.37											
5. N	36.07	25.15	-.38	-.18	-.13	-.38										
6. Self EM	2.96	.60	.32	.04	.05	.13	-.22									
7. Self SF	3.06	.54	.20	.04	.10	.12	-.16	.24								
8. Self PG	2.94	.59	.28	.11	.12	.17	-.18	.34	.20							
9. Self OM	2.93	.63	.24	.13	.05	.07	-.12	.28	.09	.40						
10. Other EM	2.67	.44	.04	.09	-.05	-.05	.07	.20	.05	.14	.14					
11. Other SF	2.96	.37	-.01	.11	.04	.05	-.01	.00	.00	.03	.13	.54				
12. Other PG	2.53	.45	.09	.08	.01	.02	.05	.13	-.02	.20	.18	.60	.52			
13. Other OM	2.31	.51	.07	.14	.00	-.03	.01	.14	-.01	.12	.25	.59	.49	.62		
14. Age	38.21	6.44	-.01	.14	-.02	-.12	-.06	.16	.07	.07	.14	.05	-.05	-.04	.12	
15. Female	.26	.44	.01	-.16	.13	.00	-.03	-.04	-.15	.03	-.02	.10	.03	.12	.09	-.21

Note: N = 378 (listwise deletion, collapsed at the leader level); C = Conscientiousness, A = Agreeableness, O = Openness, E = Extraversion, N = Neuroticism, EM = environment monitoring, SF = strategy formulation, PG = path-goal facilitation, OM = outcome monitoring. Female = 1 if female (else 0 = male). Note, given the clustered nature of the data, we do not report p-values for the significance of the correlations As mentioned in the text, the scores on the big five factors were rescaled using constants. These above means are in the original metrics.

Table 2
Results of Multivariate Structural Equation Models

Vars.	Self EM	Other EM	Self SF	Other SF	Self PG	Other PG	Self OM	Other OM
C	.27** (2.71)	-.00 (.02)	.40** (3.89)	.12+ (1.94)	.41** (3.90)	.14* (2.23)	.28** (2.82)	.05 (.70)
A	<u>-.04</u> (.76)	<u>.06*</u> (2.01)	.02 (.32)	.09** (3.05)	.06 (1.04)	.07+ (1.93)	.10+ (1.74)	.08* (2.43)
O	.13 (1.35)	.12* (2.08)	.07 (.54)	.08 (1.21)	.25* (2.18)	.08 (1.05)	.09 (.74)	.11 (1.37)
E	.11 (.81)	.08 (1.03)	.04 (.23)	.10 (1.14)	.05 (.29)	.08 (.87)	.00 (.03)	-.02 (.18)
N	-.05 (.95)	.02 (.60)	-.08 (1.44)	.06+ (1.78)	<u>-.06</u> (1.08)	<u>.05</u> (1.51)	-.04 (.60)	-.01 (.32)
Age	.00 (.35)	.00 (.89)	.02** (2.90)	.00 (.98)	.01 (1.28)	.01 (1.40)	.00 (.64)	.01* (2.25)
Fem.	-.19** (2.77)	.03 (.57)	-.01 (.13)	.10+ (1.92)	.08 (1.12)	.13** (2.61)	.03 (.41)	.12* (2.15)
Const.	4.21** (4.37)	2.50** (4.21)	4.42** (5.19)	2.62** (4.39)	3.06** (3.37)	1.83** (2.75)	1.68 (1.63)	.68 (.88)
R ²	.17	.05	.19	.07	.19	.05	.16	.07
$\chi^2(29)$	87.61**	105.27**	98.71**	162.86**	93.35**	96.67**	88.00**	167.76**

Coefficients are unstandardized; cluster robust (at leader level) z-statistics in parentheses. n = 448 leaders and n = 3,315 raters; ** p < .01, * p < .05, +p < .10. χ^2 refers to Wald test for the simultaneous test of the coefficients in the model equaling zero (i.e., R² = 0). Fixed-effects of firm, country, and time are included in all specifications. C = Conscientiousness, A = Agreeableness, O = Openness, E = Extraversion, N = Neuroticism, EM = environment monitoring, SF = strategy formulation, PG = path-goal facilitation, OM = outcome monitoring. Female = 1 if female (else 0 = male). For each pair of self- and other-ratings in the respective leader factors, (a) bolded coefficients indicated a significant difference at p < .05, and (b) underlined coefficients indicated a significant difference at p < .10. For the marginal predictions used to generate the graphs in Figure 1 we estimated the models using least squares errors-in-variables regression with clustered bootstrapping (500 replications).

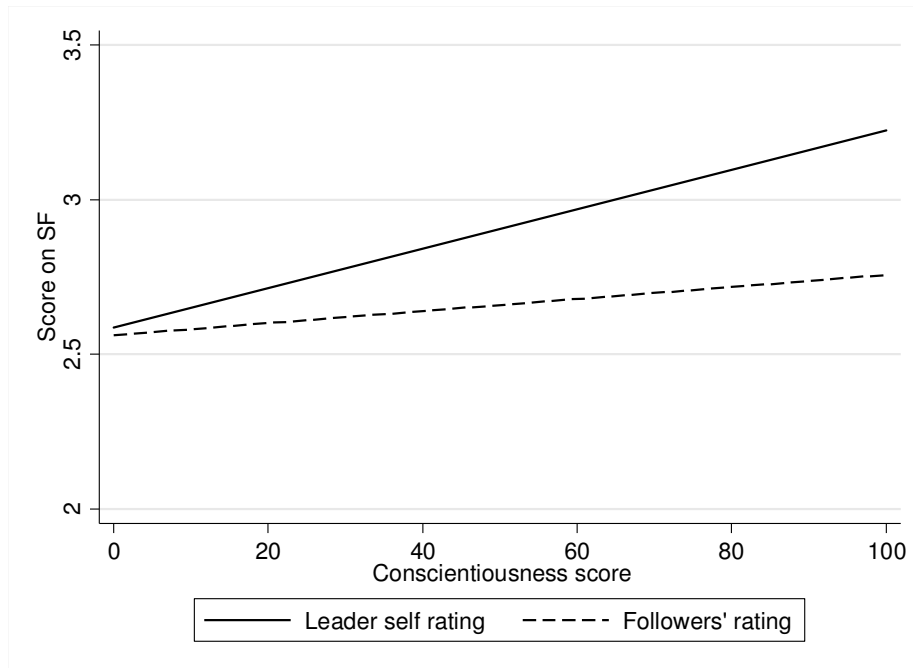
Table 3
Errors-in-Variables Results for Testing Even-Number Hypotheses

		Environmental Monitoring				Strategy Formulation			
		Margins		SOA	z- statistic	Margins		SOA	z- statistic
		Self	Other	Difference		Self	Other	Difference	
C	High	3.23	2.95	-.15***	3.83	3.22	2.76	-.44***	4.20
	Low	2.82	2.95			2.59	2.56		
A	High	3.00	3.06	.18***	4.49	3.00	2.85	.24*	2.36
	Low	3.11	2.87			2.95	2.56		
O	High	3.14	3.02	.01	.06	3.01	2.73	.03	.21
	Low	2.95	2.83			2.91	2.60		
E	High	3.10	2.98	-.03	.27	2.99	2.72	.09	.62
	Low	2.96	2.86			2.93	2.57		
N	High	2.97	2.99	.18***	3.86	2.81	2.81	.46***	8.22
	Low	3.12	2.93			3.07	2.61		
Sex	F	2.92	2.96	.14**	3.09	2.96	2.75	.11+	1.86
	M	3.12	2.94			2.98	2.65		
		Path-Goal Facilitation				Outcome Monitoring			
		Margins		SOA	z- statistic	Margins		SOA	z- statistic
		Self	Other	Difference		Self	Other	Difference	
C	High	3.19	2.63	-.37**	3.33	3.11	2.36	-.36**	3.02
	Low	2.60	2.41			2.68	2.29		
A	High	3.06	2.67	.05	.39	3.12	2.49	-.04	.37
	Low	2.89	2.45			2.82	2.22		
O	High	3.09	2.58	-.24*	1.97	2.99	2.39	.03	.26
	Low	2.73	2.47			2.86	2.23		
E	High	2.98	2.57	.05	.37	2.95	2.32	-.04	.24
	Low	2.91	2.45			2.93	2.35		
N	High	2.84	2.63	.34**	3.33	2.87	2.31	.08	.65
	Low	3.03	2.49			2.98	2.34		
Sex	F	3.02	2.63	.04	.89	2.96	2.42	.10+	1.66
	M	2.94	2.51			2.94	2.30		

Note: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$; C = Conscientiousness, A = Agreeableness, O = Openness, E = Extraversion, N = Neuroticism. Difference in agreement is the difference between SOA when a personality trait is *low* versus SOA when that personality trait is *high*. Mathematically it can be expressed as $\text{abs}[\text{Other}_{\text{low}} - \text{Self}_{\text{low}}] - \text{abs}[\text{Other}_{\text{high}} - \text{Self}_{\text{high}}]$ for any given dimension of IL. Thus, positive values signify agreement on the high end of the personality scale (or being a woman) and negative values agreement at the low end of the scale

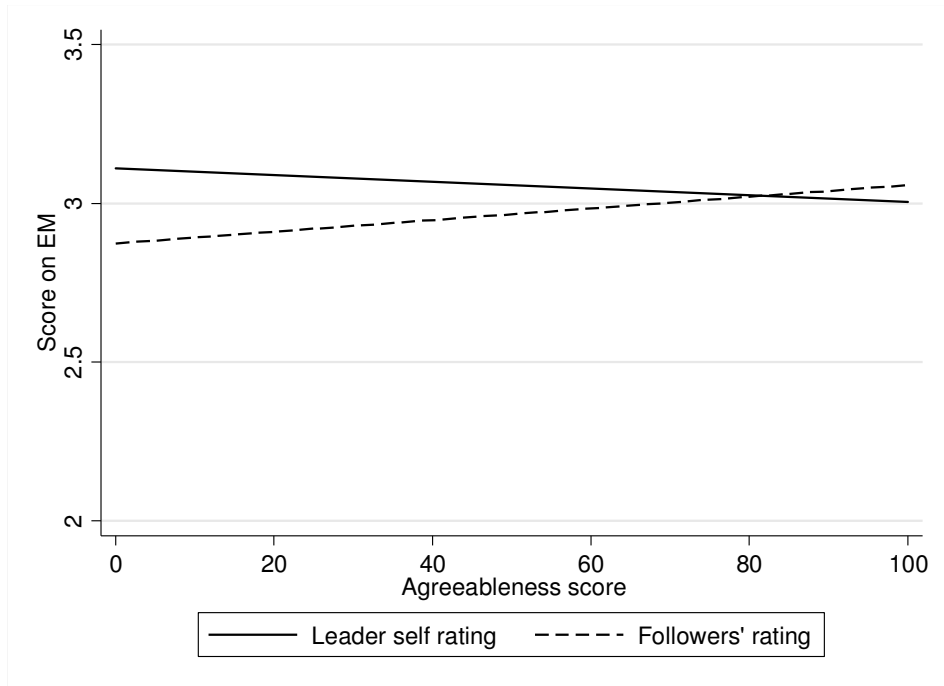
(or being a man). Cluster robust Wald tests are used to examine whether the difference is significant. A significant difference in agreement indicates that SOA varies as a function of the personality trait in question.

Figure 1
Results on the Relationship between Conscientiousness and SOA



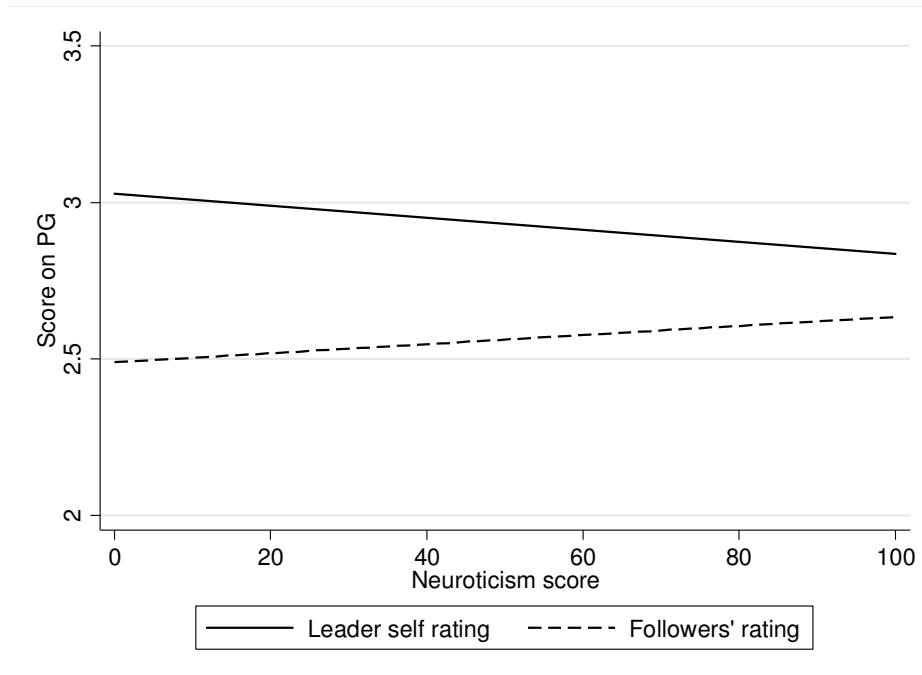
Note: SF = strategy formulation of IL; the 0 to 100 scores indicate the original scale of the measurement of personality traits.

Figure 2
Results on the Relationship between Agreeableness and SOA



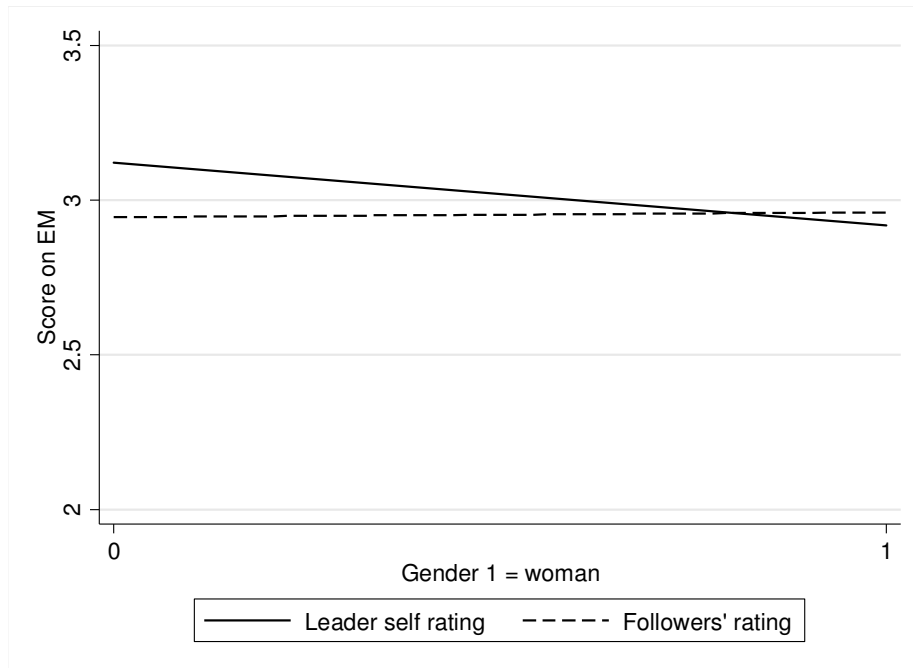
Note: EM = environment monitoring of IL; the 0 to 100 scores indicate the original scale of the measurement of personality traits.

Figure 3
Results on the Relationship between Neuroticism and SOA



Note: PG = path-goal facilitation of IL; the 0 to 100 scores indicate the original scale of the measurement of personality traits.

Figure 4
Results on the Relationship between Gender and SOA



Note: EM = environment monitoring of IL; the 0 and 1 scores on the horizontal axe indicate the dummy code for gender, where 0 = man and 1 = woman.